

Malign environmental neglect

The US Administration's indifference to environmental issues will not be removed by a few speeches at wildlife picnics.

To show that he really does want peace, President Reagan a few weeks ago made a point of praising Russian music (Tchaikovsky, to be exact). Now he is hopping around wildlife refuges to demonstrate his concern for the environment. But few are likely to be impressed by this born-again environmentalism, coming as it does so near to the elections. Indeed, what Mr Reagan may really have accomplished in his week-long carefully-planned show of concern for environmental protection is to remind the electorate of the opportunities lost by his administration to do something about some very real environmental issues.

The three and a half years of obstructionism that the administration has brought to bear on environmental policy should be seen as an opportunity lost by the administration's own supporters, not just its critics. At the last election four years ago, there was something in the view that environmental regulation had got out of hand. But instead of seizing the chance to rationalize environmental policy — particularly by instituting a more sensible approach to ordering hazards so that the response might be vaguely proportional to the risk — President Reagan, through a combination of simplistic campaign rhetoric about "big government" and a basic lack of interest in the substance of environmental issues, set the tone for the neglect that followed. The Environmental Protection Agency (EPA) — the pragmatic creation of a Republican president — was put in the hands of ideologues and time-servers who presided over its thorough demoralization. Although it would have been possible to find competent knowledgeable professionals who nonetheless shared Reagan's political philosophy, EPA (and the Department of the Interior) instead became the dumping ground for political hacks.

This approach succeeded, to be sure, in preventing new regulations (such as those covering acid rain). It also failed to do anything whatsoever about the Keystone Cops mentality inherited from previous administrations which sends EPA racing from one "disaster" to another, sometimes pushing for control of minuscule hazards (benzene in gasoline) while ignoring others a thousand times more serious (asbestos in schools).

Even after the political hackery within EPA was cleaned out with the departure of Mrs Anne Gorsuch Burford and almost all of her senior staff, obstructionism persisted. William Ruckelshaus, the new administrator, came in with a sterling reputation and a plateful of ideas for reshaping the agency's regulations along conservative but pragmatic lines. Most interesting was his concept of involving local citizens in decisions on how to balance the need for industrial emissions controls with the consequent loss of local jobs. But Ruckelshaus has found that the president's promise of his personal backing has been worth very little. Ruckelshaus was left last year with egg on his face when his efforts to strike a deal with Canada over control of acid rain was blocked by the ideologues within the administration, unconcerned about scientific details.

Symbolism has its place in government, and last week's round of visits to national parks and wildlife areas cannot be dismissed out of hand. If nothing else, it may create an expectation for substantive action that will have to be fulfilled. President Reagan will have his chance on two issues in the next few months: the reauthorization of the hazardous waste "superfund" and the possibility of reopening talks with Canada on acid rain. Failure to take a strong stand on either will surely disappoint those who have

been encouraged by recent gestures. But symbolism cuts both ways, as Mr Reagan learned from the outrage that met the appointment two weeks ago of Mrs Burford to head a national advisory committee on oceans and the atmosphere. The timing was particularly bad, coming the night before a painfully prepared meeting with four conservative environmental leaders. The meeting almost failed to take place. The four, Mr Reagan's natural allies, all heads of old-line conservation groups, most of whose members are hunters and fishermen, spent the entire meeting lambasting the President for the Burford appointment, of which Mr Reagan was apparently ignorant. Malign neglect abounds. What is needed now is some decisive action or even decisive symbolism from the top to change the tone of the administration and to permit competent professionals, such as Mr Ruckelshaus, to do their jobs. □

Miracles do not happen

A group has invited trouble by claiming that science has nothing to say about miracles.

THE presidents of the Linnean Society and of the Bible Creation Society of the United Kingdom, Dr Sam Berry and Mr E. H. Andrews, together with a vice-chancellor, a fellow of the Royal Society and other worthies, last week startled readers of the London *Times* by intervening in a theological dispute which has riven the Anglican community in Britain — the propriety of installing as Bishop of Durham a man who professes himself (on television) to have an open mind on questions such as the Virgin Birth and the Resurrection. Briefly, the Linnean president and his fellow-believers say, "it is not logically valid to use science as an argument against miracles", and "the belief that miracles cannot happen is as much an act of faith as... that they can happen".

Nobody can sensibly complain that scientists of various kinds are often religious people of one persuasion or another, or quarrel with the conclusion of Barry *et al.* that the "laws" of science are "only generalizations of our experience" and that "faith rests on other ground". But it is a travesty of something to assert that science has "nothing to say" about miracles.

Take an uncontentious miracle, such as the turning of water into wine. This is said to have happened at a wedding feast, when the supply of wine was unexpectedly exhausted. The only published account has it that jars of drinking water were found to have been transformed into wine in the socially embarrassing circumstances that had arisen. The account is now firmly a part of the Christian legend, but that is not the same as saying it is the account of a phenomenon. Obvious alternative explanations abound. As scientists, the signatories would not have given a favourable referee's opinion of such an account for a scientific journal. And far from science having "nothing to say" about miracles, the truth is quite the opposite. Miracles, which are inexplicable and irreproducible phenomena, do not occur — a definition by exclusion of the concept.

Ordinarily, the point would not be worth making. The trouble with the publication from Berry *et al.* is that it provides a licence not merely for religious belief (which, on other grounds, is unexceptionable) but for mischievous reports of all things paranormal, from ghosts to flying saucers. □



US technology controls

Curbs on conference membership advocated

Washington

FREEDOM of scientific communication with foreign nationals may be severely restricted by draft export administration regulations now being circulated within the US Government by the Department of Commerce. If the draft is approved, universities will have to obtain a validated export licence for any academic conferences attended by foreign nationals, and possibly even for undergraduate classes. The new proposals remove the existing general authorization to export educational material and information "not directly and significantly" related to industrial processes.

The Commerce Department is going to unusual lengths to keep the draft confidential. A version circulated a year ago had to be abandoned after it was widely leaked and opposed from all quarters. Since then there have been two revisions, chiefly affecting business interests, but the key elements remain unaltered.

It is unlikely that the Commerce Department's ideas will go unchallenged. A special inter-agency working group on scientific communication and export controls this week begins a "crucial" series of meetings to decide how to revise the regulations without unduly restricting scientific progress. The working group, under the chairmanship of Andrew Pettifor of the Office of Science and Technology Policy, includes representatives from the State Department, the Department of Defense and the National Science Foundation. Mr Charles Herz, general counsel of the foundation, said last week he was "hopeful" that satisfactory compromises could be found — "the aim is to ensure we don't shoot ourselves in the foot by introducing inappropriate controls". A study planned by the National Academy of Sciences of how export controls affect science will be too late to have an impact, however; the working group hopes to have reached conclusions by the end of August.

The Department of Defense recently announced that it had given up the idea of introducing a "grey area" of sensitive but unclassified research, for which scientists would be required to check with the department before publishing research results (see *Nature* 31 May, p.389). Instead, contracts would be used to impose restrictions on federally-funded research with security implications, on an all-or-nothing basis. That decision was credited to Dr Richard DeLauer, Under Secretary of Defense for Research and Engineering and a moderate on regulation.

The new Commerce Department

proposals, in contrast, are thought to owe much to the Pentagon's international security policy division, which takes a markedly different line. The use of the export administration regulations to control scientific communication, which the National Academy of Sciences two years ago considered to be the least desirable option, may lead to more far-reaching controls.

The new export regulations are being drawn up in response to a congressional mandate to include in the "commodity control list" of restricted exports a category of critical technology based on the Commerce Department's internal "militarily critical technology" list. This internal list runs to some 17 volumes and the department is apparently having some difficulty in reducing it to a manageable size and sanitizing it for publication. Although the intention of Congress in making the instruction was to avoid unnecessary restrictions on communication, there seems to be a danger it will have the opposite effect.

In parallel with the executive moves, Congress is painfully trying to revise the Export Administration Act — formally lapsed but extended under emergency powers. However, the Senate and the House of Representatives have produced incompatible bills and a conference committee has so far been unable to reconcile the two.

A requirement that universities and other educational institutions report to the government what are called "agreements to export technical data" — communication of results to foreigners — is just one area where the two houses have been unable to agree. While the House of Representatives would maintain in law the existing general export authorization for such institutions, the Senate would require reporting for technology identified by the Pentagon as "militarily critical". These reporting requirements would facilitate the regulatory changes now being aired by the Commerce Department. The conference committee will make another attempt to resolve the issue on 26 July.

Sceptics argue that the legal provisions are largely irrelevant, since it is their interpretation by the enforcement divisions that will matter. Despite the existing privileges, more and more conferences are being required to restrict access to US nationals or to restrict content. At a 1985 conference on metal matrix composites to be held by the American Society for Testing and Materials, delegates will have to exercise their ingenuity to avoid talking about "design, manufacturing, fabrication

methods, production technology or end use of the materials".

The impasse on Capitol Hill may quite possibly signal the end of the Export Administration Act. In a presidential election year, the normal rules of politics are suspended, and the act could be one casualty. Provisions to encode in statute a boycott against South Africa, pushed hard by Democrats in the House of Representatives, may be intended as much to attract a presidential veto as for humanistic reasons. The administration, however, is keen to get the act on the statute book, fearing legal challenges to the present interim provisions.

Tim Beardsley

Japanese biotechnology

Search begins for superbugs

Tokyo

A FIVE-year project is about to be launched in Japan to search the world's more bizarre environments for microorganisms that thrive under conditions of extreme pH, temperature, salinity and pressure and to try, in the long term, to use their unique properties to establish a "new biotechnology". The "Superbugs" project, as it has been officially called, is the latest addition to the innovative ERATO programme and is expected to be backed with funds of around 1,500 million yen (£4.63 million).

ERATO is run by the Science and Technology Agency's Research Development Agency (see *Nature* 305, 373; 1983) and has always sought out original — some would say downright eccentric — research themes. Six projects, including those on bioinformation transfer and biotechnology, are already under way, and a new project is being added each year.

All run under a set of rules designed to provoke unconventional enquiry: the 20–30 researchers in each project must all be under 35 years of age, they must be drawn in equal numbers from universities, government research institutes and industry to ensure diversity of background and they must work in a specially set up independent laboratory where there is no possibility that they will be contaminated by the inflexible thinking of the outside world. Or so at least the story goes. In fact a good element of hardheaded thinking goes into projects that are essentially too risky for industry to carry out but in which there is reasonable chance of long-term industrial application.

The Superbugs project leader, Professor Koki Horikoshi of the Institute of Physical and Chemical Research (RIKEN), is a pioneer in the study of microorganisms from extreme environments. During the 1970s, he discovered a whole new world of alkalophilic microorganisms, isolating several thousand strains — including even

alkalophilic phages — growing at around pH 10 (see also p. 225, this issue).

Since then, he has been studying their molecular biology and has been successful in using the new types of enzymes they produced by microorganisms in industrial processes, most notably the use of CGTase from an alkalophilic *Bacillus* species to produce cyclodextrin from starch on an industrial scale and at very high yields. More recently he has discovered a penicillase gene in an alkalophilic *Bacillus* species which, when transferred to *Escherichia coli*, changes the membrane permeability (for reasons as yet unknown), causing the bacteria to excrete penicillin into the culture medium. Commercial applications are obvious and are being followed up in Japan and abroad.

But Professor Horikoshi sees this as just a beginning. Given powerful tools provided by advances in genetic engineering, he now believes that we are falling behind in knowledge of new functions that may be usefully transferred to other organisms. In particular, industrial processes are often unsuited to the neutral pH and moderate temperatures that suit the microorganisms which have been most extensively studied so far. What better place to look for a thermostable enzyme suited to a high-temperature industrial process than in a thermophilic bacterium?

There is plenty of reason to believe that there are vast numbers of microorganisms living in extreme environments that are still waiting to be discovered. Recently a bacillus isolated from a hot spring in Yellowstone Park in the United States has been reported capable of growing at 100°C, a yeast, *Candida scotti*, has been isolated from the Antarctic which grows at 0 to -10°C but not at +15°C, and a *Thiobacillus* species growing at pH below 4 are already commonly used for leaching metals from mine wastes. And of course there is the notorious "black smoker" bacterium that grows at temperatures of around 250°C — if indeed it really exists (see *Nature* 303, 423; 1983).

What the Superbugs project now aims to do is to screen microorganisms isolated from extreme environments all over the world, analyse their tolerance mechanisms, metabolic pathways and the expression and control of tolerant enzymes, and then to try introducing activities such as alkalophilicity, halophilicity and thermophilicity into "moderate" bacteria. It may then be possible to put into operation new bioreactors that can operate at higher temperatures and salt concentrations.

Luckily, Professor Horikoshi already has a good site for the independent laboratory the project requires — the old building not far from Tokyo University that RIKEN abandoned when it moved out into the country some 15 years ago. The new RIKEN site also happily contains a new microorganism culture collection which could prove vital for storing new organisms.

Alun Anderson

US space weapons

ASAT deal in Congress

Washington

CONGRESS is heading towards a compromise that would impose a six-month moratorium on tests of antisatellite weapons (ASATs) against a target in space and further require the President to certify that a "good faith" effort is being made to negotiate "the strictest possible limitations" on such weapons "consistent with . . . national security" before proceeding with such tests.

The compromise would reconcile the differing approaches taken by the House of Representatives and the Senate to avoid an arms race in space weapons. The Senate voted by a substantial majority to impose the certification requirement only; the House, by a smaller margin, called for a one-year ban on testing against a target so long as the Soviets do not resume testing their ASATs.

The expected compromise may be approved when members return at the end of this month, and would be seen as facilitating possible negotiations between the Soviets and the United States this fall. The US Administration is at present barred by Congress from conducting tests against a target as a result of a similar amendment passed last year. That amendment required the President to certify that an effort was being made to negotiate a "ban" on ASATs. President Reagan reported last spring that a ban was impossible because of difficulties in verification.

The United States carried out a first test of its ASAT in January. No target was involved, and the rocket did not carry a warhead. Unlike the Soviet ASAT, which is fired by an inter-continental ballistic missile booster and requires two orbits to rendezvous with its target, the US system is a direct-ascent rocket fired by an F-15 fighter. The Soviet system has an unimpressive test record, with 11 failures out of 20 tests.

A second test of the US ASAT is expected any time now; it would add a warhead but again not involve an actual target. The Air Force wants to introduce a live target for the third test.

The administration, while expressing interest in the Soviet offer of ASAT negotiations in Vienna in September, has suggested that only a rather weak agreement might be possible. "Rules of the road" for satellites might be agreed to, for example. Another suggestion is a ban on tests of high-altitude ASATs, but this is not likely to go down well with the Soviets, who have a relatively greater proportion of their military satellites in low orbit. Both sides use low orbits for photographic reconnaissance and weather satellites; the Soviets in addition place a great many communications satellites in an eccentric orbit that brings them high over the Soviet Union and low over the Southern Hemisphere, and are

thus vulnerable to existing ASAT technologies. High orbits are used for navigation satellites (which are in semi-synchronous orbit) and early warning, electronic intelligence, and — in the case of the United States — communications satellites (geosynchronous orbit).

The certification requirement that the Senate has proposed would require the President not only to certify a good-faith negotiating effort but also that US tests against a target are necessary to "avert clear and irrevocable harm" to national security, that the tests will not irreversibly harm prospects for future negotiations, and that the tests are consistent with the Anti-Ballistic Missile Treaty.

Stephen Budiansky

Britons never slaves?

Los Angeles

APPROXIMATELY 40 British engineers, scientists and data-processing professionals are being brought to the United States every year by a Southern California head-hunting firm that makes employers an offer that is hard to refuse.

The British professionals agree to work for less money than US counterparts, are "highly motivated and expertly trained", and once in the United States, cannot change jobs. Work visas are good for three years, whereupon the British imports must return home.

The practice has infuriated a group called the Committee of Concerned Electrical Engineers which compares the head-hunters with the slave traders of 200 years ago.

According to John Alcorn, co-founder of International Staffing Consultants (ISC) of Newport Beach, California, there are genuine shortages in the United States of systems analysts, programmers and computer engineers as well as "high-tech engineers in speciality areas". Two jobs recently filled by Britons were as a ring laser gyroscope engineer and an electrical discharge machine operator.

The US Labor Department protects American workers, Mr Alcorn said, by demanding that employers advertise jobs before letting foreigners apply. But this requirement can be bypassed if foreigners are hired on temporary visas.

ISC, which serves "a couple of dozen companies", ungrammatically advertises that "The opportunity to work in the United States is a powerful motivator for the British DP professional. They earn approximately one half of the salaries paid to US employees, so working for you is viewed as an exceptional opportunity. Unlike US employees, you won't risk losing them to your competitors."

Sandra Blakeslee

US nuclear waste

Widespread problem of disposal

Boston

TIME is running out for individual states in the United States to develop facilities for disposal of low-level radioactive waste produced by nuclear reactors, industry and biomedical research and treatment. The federal Low-Level Waste Policy Act of 1980, pushed through Congress in the waning days of the Carter Administration, required the states to form regional compacts to plan and develop dumps for the low-level material generated within their borders. The deadline is January 1986, just eighteen months away.

There are at present only two sites operating, in the states of Washington and South Carolina, which take most of the low-level waste arising in the United States. After the deadline, those states will have the right to refuse waste from outside the compacts they have formed with neighbours. Political pressure has only recently forced other state governments to plan more decisively. But since the planning, public approval, development and construction of a dump will take at least five years, states outside the South Carolina and Washington compacts are unlikely to meet the deadline.

In Massachusetts a Special Legislative Commission has written a working draft of a compact that would be acceptable to the state, and is now working out guidelines for choosing potential sites.

But Massachusetts has special problems. Its strong environmental and anti-nuclear constituencies will help to ensure that the site is safe and well-contained, but a referendum proposed by activist groups and approved by state voters in 1982 will inevitably slow down the approval process. The law, known as Chapter 503, requires that once a developer has found a site, obtained all the necessary licences, performed environmental impact studies and site characterization (at a cost of at least \$6 million), both houses of the state legislature must vote that the technology and the site are superior to all others, whereupon the proposed facility must be approved by the voters in a statewide election, the next of which is in November 1986 — eleven months after the deadline. Moreover, the other New England states have understandable doubts about whether Massachusetts could ever fulfil its part of a compact.

The potential restrictions on waste disposal after 1986 could drastically affect work that produces low-level waste unless the grace period is extended. Although the amount of waste generated has been greatly reduced (by two-thirds in the past five years in Massachusetts), some interim arrangements will have to be made. The four largest producers of radioactive waste may be able to store their own waste until a site is developed, assuming they will no

longer have access to the existing sites.

Small producers, such as hospitals and universities, are hoping to avoid the huge expense of storing radioactive waste. Partly in an effort to economize, they have already made substantial reductions in their waste production, and are storing rapidly-decaying elements such as ^{32}P through fifteen or so half-lives. Any plan to build or convert a warehouse for storage is likely to face strong local opposition — and research institutions in Massachusetts are already in bad odour. Much biological waste is decomposable, making temporary storage impossible and leaving only the undesirable alternative of incineration. Although large producers such as NEN are already preparing for the possibility that they will have to store their wastes, hospitals and research institutions appear

to have no strategy to fall back on.

The small producers are hoping the state will provide an interim storage facility. The Massachusetts Special Legislative Commission intends, however, to concern itself only with the long-range solution. The state government hopes the federal government will assist the process by extending the grace period and improving the laws, regulations and technology that the states must follow in creating new waste dumps.

And everyone is hoping that the states of Washington and South Carolina will cooperate and continue to allow low-level waste from other states into their dumps after the 1986 deadline. Whether they will agree is doubtful — there is strong feeling within the states that there should be no outside access. South Carolina, for example, has threatened that if Congress extends the deadline stipulated by the Low-Level Waste act, it will shut down its facility entirely. **Christopher Earl**

Dispute over AIDS patent priority

Washington

WHILE the public health problem posed by acquired immune deficiency syndrome (AIDS) is causing growing concern, industrial companies are not losing sight of the commercial possibilities. Genetic Systems Corporation (GS), of Seattle, Washington, has formed a joint venture with Pasteur Institute Productions of Paris to develop and market a diagnostic blood test for lymphadenopathy virus (LAV), the AIDS-associated virus observed by Dr Luc Montagnier, and claims patent priority for the venture over the test developed by Dr Robert Gallo for the (probably identical) human T-cell leukaemia virus III (HTLV-III). GS has been assigned exclusive rights to market the test in the United States and Canada, while Pasteur will market the test in Europe. To complicate the issue, however, Biotech Research Laboratories, of Rockville, Maryland, plans to challenge GS's claim, saying it has an even earlier patent for a general HTLV diagnostic test that has been on sale for over a year and will now also detect HTLV-III.

Biotech was supplied with Gallo's first HTLV isolate to develop an enzyme-linked immunosorbent assay (ELISA) and filed for a general patent on the technique in the summer of 1983. The company is backed by Du Pont, one of the five selected by the federal government to produce Gallo's test (see *Nature* 5 July, p.6) and is contracted to undertake its manufacture. Dr Robert Ting, director of Biotech, is confident that the HTLV test he patented will cover the test for HTLV-III, even though this was discovered later. He will not use his patent rights to oppose the federal government's chosen five companies but would move swiftly against GS. The stakes are high: the world market for an AIDS blood test is put at not less than \$100 million a year.

The basis for the GS claim is that the Institut Pasteur in Paris filed for worldwide patent rights on its LAV test in September 1983, while Gallo's application, covering his high-yield permissive cell line, the uninfected cell line and the specific ELISA test for HTLV-III, was not filed until May this year. Dr Robert Nowinski of GS says it is "generally accepted that the French have priority of inventorship".

Gallo says he is "exasperated and totally fed up" with patent priority disputes. He argues that a useful blood test for HTLV-III was impossible before his discovery of the HT permissive cell line; the method for production of LAV recently described by the French group (*Science* 225, 63; 1984) is, he says, nowhere near as useful. Furthermore, the tests used by the two groups are fundamentally different. By looking for antibodies to the p41 envelope protein of the virus, Gallo says he can identify AIDS patients with 100 per cent efficiency. The French group, in contrast, have not so far been able to detect this protein, and antibodies to core proteins are less well correlated with disease symptoms. Gallo says AIDS is too serious to waste time arguing about priorities: his virus and sera are freely available to any qualified recipient, and indeed reagents were supplied to the French group during their early investigations. Stocks have also been sent to Britain for use there in blood bank testing.

All parties agree that the squabble over patent priorities should not hinder the most rapid possible availability of a test. Dr Montagnier in Paris says his patent application would not be used to prevent others from producing a test; GS agrees but wants "proper allocation of commercial rights". But Gallo says of GS that "they're only in it for the money". **Tim Beardsley**

Atmospheric science

UK initiatives for change

THE problems of acidity and ozone in the lower atmosphere have thrown into relief urgent questions that need to be answered by atmospheric chemists. At a meeting in Oxford last week, convened on the initiative of the Natural Environment Research Council (NERC), the UK atmospheric chemical community gathered to survey some of the fundamental aspects of their discipline that require investigation, in the framework of which the more politically charged problem of pollution can be tackled.

The meeting represented the first public result of a decision by NERC about a year ago to designate atmospheric chemistry a "special topic". The Coordinating Committee for Research into Atmospheric Chemistry (CCRAC) was formed and £250,000 was set aside for a three-year programme to stimulate research in an area in which UK expertise was somewhat fragmentary.

That the UK community is still fragmentary cannot be denied — although, as one overseas visitor commented, many of the fragments are of a sufficiently high calibre for fruitful international collaborations, collaborations that are particularly necessary in a field where global moni-

toring of species' behaviour, for example, is a basic requisite for effective research. But the organizers of the meeting were encouraged by the large attendance, including some for whom either the atmosphere or chemistry, but not both, had previously figured in their grant proposals. CCRAC's hope is that such people will be encouraged further into the field, partly by the availability of specially directed grants.

International interest in this field is sharpened by the imminent publication by the US National Academy of Sciences of a detailed survey of global tropospheric chemistry. If a large US programme eventually results from this report, then the UK and other European research communities might benefit. If the UK position looks bleak, on the other hand, researchers may be tempted to emigrate.

CCRAC includes representatives of institutions such as the Science and Engineering Research Council, the UK Atomic Energy Authority and the Meteorological Office. An interesting example of the interplay of chemistry with other atmospheric phenomena was provided by Dr A. Tuck of the Meteorological Office, who demonstrated the mixing of tropospheric and stratospheric air — with corresponding effects on the ozone distribution — apparently confined by a low-pressure system off south-west England, the mixing region coinciding with the low-pressure centre. **Philip Campbell**

UK higher education

Student demand increased

THE UK Committee of Vice-Chancellors and Principals has moved swiftly to take advantage of the British Government's revision of its estimates of future demand, published on 12 July. In a statement issued last week, the committee sharply argues that since the demand for higher education will not fall, for demographic reasons, below the present number of full-time students, final calculations of the provision of student places should be postponed until later in the decade — and that budget cuts now threatened should be abandoned.

The affair is bound to be an embarrassment for the Department of Education and Science, which has during the past four years sought to justify the budget cut of 8.5 per cent announced at the end of 1980 on calculations of reduced demand for places in higher education, itself a consequence of the reduced birth-rate in the late 1960s.

The paper published a year ago by the Royal Society seems to have been especially influential in causing the revised estimate, chiefly because of its explicit use of data on social mobility, and the educational attainments of women students. The vice-chancellors said last week, however, that even the more recent government projections did not take full account of upward mobility, pointing out that the rapidly increasing proportion of parents who are graduates may be more relevant than social classes as conventionally defined.

The vice-chancellors also complain that the new projections do not make sufficiently generous allowance for the expected increase of demand from mature students (over 26) or for the expected attainment of women school-leavers.

The statement says that financial provision for higher education in the next three years is now "considerably out of line" with the government's new estimate of demand and, ominously, promises further study of the issue.

Other developments in the past week include:

- Moves supported by the Foreign Commonwealth Office to abate the cost of fees for overseas students in British universities.

- A bid by the Science and Engineering Research Council for funds to collaborate in the European Synchrotron Project (disclosed to a House of Commons select committee).

- A lobby of last week's meeting of the Medical Research Council by members of the Association of University Teachers (and other groups).

- A mounting sense of despair among research council heads, one of whom said this week that it "would not be time to take to the streets" until November. □

New Polish university proposed

THE Polish Government has placed a bill before the Sejm (Parliament) authorizing the establishment of a university in Szczecin, a surprising development at a time of severe economic restraint and cutbacks in education (about 50 university-level courses have disappeared during the past three years). According to Warsaw radio, however, the university is needed for "state and social reasons", including the "traditions of Polish culture" in the north-west of the country and the "requirements of the state's maritime policy".

Although Szczecin is one of the few major Polish cities to lack a university, it does have a number of higher educational institutions, including a polytechnic, a teachers' training college, an agricultural academy, a medical academy and a higher maritime school. The agricultural academy includes a flourishing department of marine fisheries and food technology, with institutes of aquaculture, fisheries' economics, ichthyology, oceanography and conservation, and sea-food technology, which, together with the higher maritime school (which specializes in shipbuilding and engineering) would at first glance seem sufficient to satisfy the country's requirements. (Indeed, since the agricultural academy also runs diploma and PhD courses for foreigners in fisheries-

related subjects, it might be argued that the academy has, in fact, a teaching capacity in excess of those requirements.)

The "traditions of Polish culture" in the area may constitute a more cogent argument. Before 1945, Szczecin had been under German rule for three centuries, as part of the Electorate of Brandenburg. The frontier changes after the Second World War meant that the German population of the territories reacquired by Poland migrated westwards, mostly to the Federal Republic of Germany. Of recent years, some of these elderly exiles have been sending letters to the Polish inhabitants of what they still consider to be their homes, warning them not to put down roots, since sooner or later the Germans will be back.

The West German Government is embarrassed by the whole affair and has disclaimed any responsibility for what it stresses are the individual views of only a tiny minority of its citizens. The Polish authorities fear, however, that such a campaign could not be launched without the tacit approval of some person or persons high in government office. Any proposal that will reinforce the Polish claim on the "western territories", such as the proposed university in Szczecin, is a strong candidate for funds even in the current economic difficulties. **Vera Rich**

Israeli universities

Faculty forced into salary cut

Rehovot

WHEN every occupational group in Israel is wildly scrambling for wage increases, Israeli professors and senior lecturers have agreed to a 6 per cent cut in wages. "We took the step", says Professor Menachem Magidor, chairman of the Coordinating Council of Faculty Organizations, "in order to prevent a generation of young scientists from being wiped out."

Behind the melodramatic language lies the brutal fact that Israeli institutions of higher learning, now in the midst of a severe and prolonged financial crisis, have had to dismiss more than 200 members of their academic staffs over the past year and will have to dismiss more unless some solution to their budgetary problem is found.

The 6 per cent pay cut in itself is significant only insofar as it may prod the government into increasing its support for the universities. Indeed, shortly after the professors made their statement, the government said it would cover some of the \$50 million debt accumulated by institutions of higher learning. But a pledge made shortly before election day may have little relevance after polling has taken place.

Nobody doubts, in any case, that Israeli universities and research centres are in trouble and that standards are falling. In the recent report of the planning and grants committee of the Council for Higher Education, its chairman, Professor Haim Harari, pointed out that during the past decade the student body of local universities has risen by 30 per cent while the teaching staff has fallen by 3 per cent, and that there are serious problems with equipment and facilities.

"What is at stake", Harari declared, "is not merely the future of higher education and scientific research but the future of Israel itself. Israel has no chance of developing and prospering if it is not in the front ranks of science and culture."

The sophistication of Israel's industry and agriculture, he says, reflects the achievements of Israeli institutions of higher learning 10 or 15 years ago. Today, in contrast, "the people we will need in the 1990s are being trained in far inferior conditions". And there is a great difference between a young scientist who gets personal attention from his teachers, and who uses advanced equipment, and one who is trained at a university where the equipment is old and in poor repair, and where the ratio between students and teachers is more appropriate for a high school than for a university.

Research is also hampered by financial constraints, even where it has clear economic importance. One researcher working in the field of artificial intelligence

has pointed out that people at Massachusetts Institute of Technology working on artificial intelligence and related subjects have at their disposal about 50 VAX computers and 50 others, while he has only one of each.

This situation is not likely to encourage other young Israelis, now holding top positions in the United States, to come home. Some now working at such places as IBM, Bell Laboratories and Harvard University have recently made it clear that they would be ready to return immediately if only they could be assured of the equipment they need to carry out their experiments. But the money simply is not available.

Some researchers find themselves in another kind of Catch-22 situation. A leading Israeli expert on applied physics has been offered more research grants than he can possibly handle. But his institution has put a freeze on hiring, and institutional salaries are, in any case, so much lower

than those offered by industry that he is unlikely to find qualified engineers and technicians for his research teams.

Most institutions of higher learning have tried to solve their problems by cost cutting, but the Haifa Technion has adopted the opposite approach. Its international board of governors has decided to ignore the 10 per cent cut urged by the government and to increase its budget for the next academic year by \$4 million so as to purchase new equipment and to hire 165 junior faculty members for its computer sciences, aeronautical engineering, industrial engineering and mechanical engineering faculties which, it says, should be enlarged to meet the country's needs.

The Technion obviously hopes that the government, impressed by the need to train manpower for science-based industry, will give it more money, even if this means giving less to other institutions. It remains to be seen whether this hope has any basis in reality.

Nechemia Meyers

Rows continue over US patent life

Washington

BIOTECHNOLOGY companies are complaining that they have been left in the lurch by a compromise between two major camps of pharmaceutical companies on extending the life of patents.

The major research-intensive drug companies have for years been asking Congress to "restore" to their patents the time lost in meeting government safety-test requirements. According to the Pharmaceutical Manufacturers Association, 8.5 years of the 17-year life of a patent is lost in carrying out the required tests and obtaining approval from the Food and Drug Administration (FDA) to market new drugs.

Legislation that would extend patents to compensate for regulatory delays has met stiff opposition from another group of drug companies, the so-called "generic" manufacturers whose principal business is making Brand X versions of patented drugs after the original patent expires. The generic versions are always considerably cheaper than their brand-name competitors, a fact that has never pleased the traditional companies. The generic companies maintain that the public interest is best served, however, by encouraging open competition as soon as possible.

Under the recently agreed compromise, now before Congress (HR 3605), the traditional companies would get their patent extension (one-half the time they spend testing plus all of the time FDA spends processing the application — up to a maximum of five years' extension); in return, the generic companies would be permitted to file "Abbreviated New Drug Applications" (ANDA) for their look-alike drugs.

These applications would permit the generic manufacturer to use the safety and effectiveness data already submitted to

FDA by the original manufacturer. Under the present rules, generic manufacturers must duplicate those tests unless they have been published in the open literature. FDA officials have said for some time that they consider this to be a wasteful process, and one that results in unnecessary clinical trials on humans.

A much more controversial provision of the compromise allows the generic companies to jump the gun — to begin testing the original manufacturer's drug even before the original patent expires, so as to be in a position to hit the streets as soon as the patent expires. (The present law forbids any commercial use by a competitor until the patent expires.) The proposal would also allow a generic company (or anyone else) to challenge the validity of a drug patent at any time simply by filing an abbreviated new drug application for the same product.

Ronald Cape, chairman of Cetus, testified before a House of Representatives subcommittee last month that the compromise — while perfectly sensible for major drug companies that have a continual stream of new products going onto the market — would inadvertently harm the emerging biotechnology industry. The prospect of extended patent life means relatively little to the biotechnology companies — which are only now getting their first patents — and hardly compensates for the prospect of immediate harassment that the ANDA challenges may bring. "A very large portion of our resources could be tied up fighting these ANDAs", he said, which would encourage "premature litigation" — especially harmful to a young field such as biotechnology which has not even had the benefit of judicial determinations of the validity and scope of its patents. Stephen Budiansky

Chinese census

Demographic transition under way

Washington

A NEW analysis by the US National Academy of Sciences* of recently released census data from the People's Republic of China has confirmed that there were some 27 million excess deaths during the "Great Leap Forward" in 1960, and provided indications that female infanticide may still be taking place in rural areas.

The data also show the substantial progress China has made in reducing both birth and death rates. The average lifespan, calculated from the new 1982 census data, is 66 years for men and 69 years for women, roughly comparable with that of Western countries in 1950. In the 1950s in China, the average lifespan was almost 25 years shorter — roughly comparable with the United States in 1890.

The birth rate in China has fallen from 40 per 1,000 in the 1950s and mid-1960s to half as much; expressed another way, women are now bearing children at an average of 2.6 per lifetime, down from 6.

Until last year, the Chinese Government had released only the barest minimum of census data. Censuses were conducted in 1953, 1964 and 1982, the latest being the most comprehensive and the first for which the Chinese sought outside assistance. Details of all three censuses were published last year in a Chinese journal; for the first time, the Chinese revealed data on its population distribution by age and sex. In addition, results of a fertility survey of a million Chinese women, conducted in 1982, have been made available; these results include marriage and birth rates broken down according to age of women for each year from 1950 to 1981.

According to Ansley Coale of Princeton University, who prepared the academy report, the data show a consistent under-reporting of births and deaths in the official registers. The census data themselves, on the other hand, show a remarkable internal consistency; especially striking is the almost perfect match between the total population figures from the census and the expected increase in population as calculated from the fertility survey and census-derived death rates.

The data on the "Great Leap Forward" — Mao's attempt at industrial and agricultural modernization that is said to have led to a near-collapse of the food distribution system in China — show sudden increases in the death rate followed by a sudden drop in fertility rate. Coale said the fertility decline was characteristic of a "quasi-biological" cause, rather than deliberate contraception; the drop occurred in women of all age groups.

The analysis of female births suggests that 60,000 female infants are "missing" each year. The normal ratio of males to females at birth is 106 to 100. First and second births in rural areas are very close to

that ratio; but for third and subsequent births there are 113 males to every 100 females. Coale says that sex-selective abortion is not a plausible explanation in rural areas; he also noted the regular denunciations in the Chinese press of female infanticide as a "feudal" practice. It is also possible, he said, that parents may be willing to face the penalties that come with having more than one child if the child is a boy, but are less likely to report the over-quota births if the child is female.

Coale's analysis also explains the recent increase in birth rate, seemingly breaking the trend of the past decade. Coale found

that official pressures to delay marriage had in fact played a significant part in the declining birth rate; a constantly increasing marriage age effectively buys time by postponing first conceptions by a new generation by women of child-bearing age. The recent decision by the government to relax the unpopular policy of delayed marriage has sent the fertility rate back up. On the other hand, the data do show substantial acceptance of contraceptive practices, with the age-specific fertility rate following closely the pattern seen in Western countries — much lower fertility among older women, who have already had all the children they want. **Stephen Budiansky**

**Rapid population change in China, 1952-1982, National Academy Press, Washington, DC, 1984 (\$12.50).*

Learning Japanese

Plans for language teaching

Tokyo

IN line with its new status as the world's number two industrial power, Japan is now laying plans to become one of the world's leading educators. At present, only 10,428 foreign students are enrolled in Japan's universities — compared with 310,000 in the United States, 119,000 in France, 57,000 in West Germany and 52,000 in the United Kingdom. But a new report predicts that Japan will be in third place by the end of the century with 100,000 foreign students.

The report, from an 18-member committee of the Ministry of Education, Culture and Science headed by Shigeto Kawano, director-general of the Japan International Education Association, proposes fundamental reforms necessary to internationalize Japanese education and the Japanese language.

First and foremost, there has to be an improvement in the availability and standard of Japanese language education. Around 1.4 million people outside Japan are now thought to be studying Japanese, one million of them in China and the remainder in the countries bordering the Pacific Ocean, particularly South-East Asia and Australia.

Within Japan, there are some 26,000 foreigners studying Japanese, three times as many as ten years ago. But, serving these foreigners, there are only around 1,000 Japanese teachers and in the whole world there are only 2,400 institutions that can offer any kind of Japanese language course. Japan has no equivalent of the British and US-supported programmes for teaching English as a foreign language, no set of standard examinations by which foreign students can measure their progress or gain official recognition of their proficiency, no authorized textbooks and virtually no textbooks at all for those who wish to learn advanced colloquial Japanese.

To help end this chaotic situation, the report calls for the number of teachers in-

side Japan to be increased to 4,200 by 1992 and to 10,600 by the end of the century. A separate report from the ministry's Agency of Cultural Affairs suggests the establishment of a four-level standard proficiency examination to be held at at least twenty centres throughout China, South-East Asia and Australia.

Next, there will have to be reforms to make it easier to enter Japanese universities including the mutual recognition of academic qualifications and the setting up of a unified entrance examination suitable for all universities that can be taken abroad, the establishment of special short-term courses, changes in university timetables and an increase in accommodation for foreign students.

The plan will also require many more scholarships to be made available. Of the students at present in Japan, only 2,082 are sponsored by the Ministry of Education, Culture and Science. The report asks that scholarships should support 10,000 out of the 100,000 students there will be by the year 2000.

A big problem, however, is that as student numbers increase, more and more of them will be expected to attend private universities. Japanese private education now costs at least ¥1.3 million (£4,000) a year, and, as many of the students are expected to come from less affluent Asian countries, there is the risk that Japan will price itself out of the education market as its inexorable economic growth continues.

Alun Anderson

Correction

OUR Israel correspondent, Nechemia Meyers, says that he was mistaken in saying (*Nature* 3 May, p.8) that Dr George Kanazi's work on mediaeval manuscripts on the art of wine-drinking had been described in a lecture at the Israeli Academic Centre in Cairo; rather, an extensive account of the work appeared in the centre's *Bulletin*. □

Of plants and men

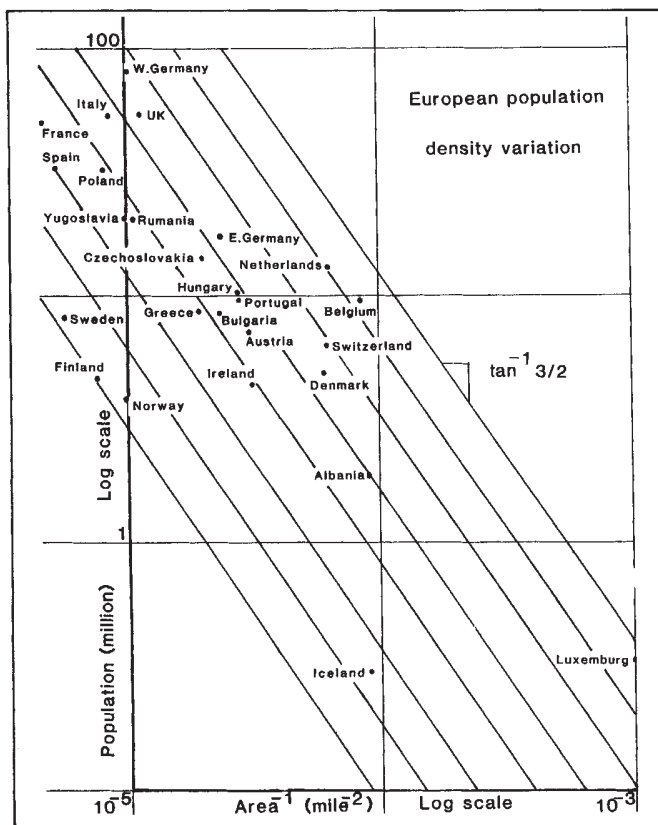
SIR — Your review of science in the Low Countries (*Nature* 7 June, pp.491–510) emphasized the contrasts of administrative style and professional practice between these ostensibly similar countries.

It was, however, the issue of overall similarity that reminded me of an earlier casual observation I had made of a human parallel to the so-called “ $-3/2$ ecological power law” of plant distribution. In essence, this law — in fact, a rule — describes the relationship between numbers of a given species in a given area and their combined biomass in such a way that a graph of biomass against distribution density, both on a logarithmic scale, exhibits a linear relationship with a gradient of $-3/2$. This observation seems to have originated from Japan^{1–3} and has yet to be properly explained⁴ for plants.

For my part, as an ecological novice, I attempted to explore the possibility that the “ $-3/2$ law” may be represented in faunal distributions as well as plants. Unfortunately, animals are more mobile than plants. Also, the animals for which most distribution data are available to the non-specialist — man — live in civilized communities within which false distributions arise due to activity specializations. Therefore, I decided to test the theory at a national level by “weighing” whole nations, that is, noting the population size, and simultaneously determining their distribution density as 1 nation per national area, that is, $1/(national\ area)$. This removes the anomalies arising from urban centralization, etc.

The outcome of this trivial exercise is most enlightening. On a worldwide basis, the aforementioned log-log graph appears at first sight to be a scatter plot. However, if straight lines with a gradient of $-3/2$ are superimposed on it, the nations line up in well ordered fashion, with those which might instinctively be thought to be similar appearing on the same line as one another and distinctly separate from dissimilar nations. A particular case in point is represented by Luxembourg, Belgium and the

Netherlands, the population densities of which, in that order, appear to be similar to the secular development of a group of plants of one species. One could go on to draw attention to certain striking anomalies and to the historical, even anthropological, comments which are thus



available to any reader prepared to repeat the diagram from simple demographic publications such as a family encyclopaedia. However, suffice it to note here that these observations are made entirely empirically — in the spirit of the original “ $-3/2$ law”. Nevertheless, their implications may not be lost on those familiar with the all-embracing Gaia⁵ concept.

A.F. HAYTON

18 Thorp Cottages,
Ryton, Tyne & Wear, UK

1. Shinozaki, K. & Kira, T. *J. Inst. Polytech. Osaka City Univ.* D7, 35 (1956).
2. Tadaki, Y. & Shidei, T. *J. Jap. For. Soc.* 41, 341 (1959).
3. Yoda, K., Kira, T., Ogawa, H. & Hozumi, K. *J. Biol. Osaka City Univ.* 14, 107 (1963).
4. White, J. *J. theor. Biol.* 89, 475–500 (1981).
5. Lovelock, J.E. *Gaia: a New Look at Life on Earth* (Oxford University Press, 1979).

Selfish ideologies

SIR — Surely your opinion of *Nature*'s readership cannot be so low that, in order to hold our interest in a discussion of the bankrupting of Third World nations, you have to convince us that this would “[impede] the pursuit of research objectives of all kinds” (“Planning for next

year's summit”, *Nature* 309, 569; 1984). You might equally well challenge us to oppose nuclear warfare on the grounds that it would fog our autoradiographs.

At a time when the dominant political ideology encourages each group to defend its own interests with little regard to the suffering of others, it might be hoped that the scientific community, which purports to cherish ideals of vision and cooperation, would adopt a less parochial mentality.

Let us be honest; the World Economic Summit achieved nothing for the unemployed in the West or for those in poverty in the Third World. Of course we need growth and innovation, but not to prop a tottering economic system. Rather, we need programmes of technological development and production designed to meet the real needs of our populations. Such programmes will not be forthcoming until the selfish ideologies reflected in your editorial are rejected.

P.M. BRICKELL
D. MURPHY

Department of Biochemistry,
Imperial College of Science
and Technology,
London SW7 2AZ, UK

Mrs Malaprop at work

SIR — It was interesting to see the article on molecular conformations in surfactant micelles (*Nature* 3 May, p.42) listed in the contents page as “molecular confrontation”. True, there has been confrontation over the issue of micelle structure, but it has been between men and not molecules. To stretch it down to the molecular level is being a little anthropocentric, is it not?

D. BALASUBRAMANIAN

Centre for Cellular and
Molecular Biology,
Hyderabad 500007, India

Trouble whipping up

SIR — As a one-time farmer but now dependent on commercial eggs for my cooking, I suggest a crucial control is missing from the letter from McGee *et al.* (“Why whip egg whites in copper bowls?”, *Nature* 308, 667; 1984). Their eggs appear to come from a commercial supplier and it is unclear whether or not they were free range. In my experience, the whites of fresh, free-range eggs beat more easily than do “fresh” commercially supplied ones, which may actually be several weeks old (they also make better mayonnaise, but that is another story).

The diet of free-range hens may differ in the availability of trace elements such as copper, or the macromolecular composition of the egg white may change with ageing, or both. Can I suggest McGee *et al.* repeat their experiment with eggs of more accurately defined provenance?

The Open University,
Walton Hall,
Milton Keynes MK7 6AA, UK

STEVEN ROSE

Parental consent over embryos

R.G. Edwards & M. Puxon QC*

A British researcher and a barrister argue the need for clarification of the role of genetic parents in the determination of the use made of fertilized embryos in research and medical practice.

THE problems surrounding research on spare embryos are new to ethics, and strike at the root of life itself. On the one hand, such research could open undreamt-of vistas of improvement and amelioration for mankind; on the other, it raises fears that, by interfering with life's earliest beginnings, it may threaten the sanctity of human life in general. The kernel of the legal problem is this: when does an embryo become a human being, and therefore attract the protection given by the law to helpless members of society? The ethical problem may well be different, and might depend on the answer given by the law to the essential question.

At present there is no legal definition of the beginning of life. It is illegal to terminate a pregnancy¹ except under certain conditions related to the health and welfare of the mother², but there can be no abortion where there is no child. Recently, in the furore over the post-coital pill, the Attorney-General refused to prosecute those who prescribed the pill on the basis that there could be no pregnancy where the fertilized ovum had not imbedded: by analogy, the embryo produced *in vitro* is not a life which demands the protection of the law.

How does this affect the requirement for parental consent for research on such embryos? Those organizations which concern themselves with the ethics of human fertilization *in vitro* insist that such consent should be obtained before research is carried out on the embryos growing in culture. A limit is generally prescribed for the duration of embryo culture — for example, 14 days after fertilization — even by those who recognize the desirability of research. The question of whether or not parental consent is necessary cannot be decided by reference to the practice universally followed by physicians and institutions, because a different entity is under consideration.

Nevertheless, consent would seem to be desirable, even if not strictly necessary for research, whatever the legal status of the embryo may be. After all, the gametes come from the parents and must "belong" to them in some sense of the word. But consent is not a clear-cut issue, and before anybody sets out to obtain such consent, some important questions should be asked

about the motives for obtaining that consent. The first motive inevitably is to avoid legal complications. Where the parents give their consent freely, with clear knowledge of what the programme of research entails at least in general terms, they cannot thereafter complain, let alone take legal action against the researcher or the institution where it is carried out. Ethical considerations apart, this is an obviously sensible step to take.

But there is another important aspect of parental consent. The parents have a responsibility for their offspring, and it may well be argued that this responsibility extends to the early embryo. What the parents' rights and duties are has not yet been decided by the law, but on the assumption that the embryo "belongs" to its parents, as a sort of chattel, and is their sole property, they can dispose of their embryos as they wish.

Many people would insist that this assumption is wrong, and that embryos have rights of their own, even full human rights, immediately upon fertilization; for such people, parental consent would not overcome the affront to the embryo. If it is argued that the consent of the embryo is required, which of course it cannot give itself, then on the analogy of the young child giving consent to an operation, its parents could consent on its behalf. Neither an institution nor a professional organization, nor even a government, could give such consent, and any interested party could step in if such consent were given by the parents and make the child a "ward of court"; and the court would probably refuse to give consent on its behalf.

One thing is certain: no form of consent by the parents can exclude the rights of the child, when born, to sue the scientist or doctor for any damage suffered as a result of negligence in culture maintenance, research or freezing at the embryonic stage³. While a live birth following embryo research may seem unlikely at present, with the application of new techniques of sexing and identification of genetic defects, such a possibility must be faced.

The question of consent for embryo research is particularly important in cases of known genetic disorders in the patients, husband or wife, to avoid replacement of a genetically disabled embryo. The question becomes even more poignant in cases of embryo donation and surrogate motherhood.

If research on embryos is to be generally accepted and carried out, it can be done only on the basis of two assumptions. First, embryos *in vitro* do not have any rights until a stage of development well past the proposed studies; in this case, parental consent would be enough. Second, any request for consent from the parents inevitably places them in a predicament. In the case of a developed child, parents would in general be acting wrongly if they gave their consent to any procedure which was not to the direct advantage of the child, except in unusual circumstances such as kidney donation between identical twins. The situation concerning parental consent arises poignantly with fetuses as Teifel³ has pointed out: parents aborting their fetuses have condemned them already, and are the last people to be asked to give consent for research, for ethically it can be seen as a double wrong to the child.

If this argument about advantage applies to embryos, the only parental consent possible would be to replace the embryo in a recipient to confer its best chance of survival. But to most of us, there is a great ethical difference between a minute cleaving embryo and a fully-differentiated mid-term fetus, so that the analogy is not a helpful one. The legal clarification of embryonic rights, if any, is a first step to the solution of this dilemma.

The need for research is fundamental to the process of *in vitro* fertilization. Scientists and physicians have a duty to ensure that an embryo replaced in its mother for growth to full term is as normal and healthy as it can be. Their duty to a potentially live-born child must be overwhelmingly greater than to the cleaving embryo *in vitro*, and this duty must apply to attempts both to alleviate infertility and to avert inherited and other defects in the resulting children⁴. To undertake *in vitro* fertilization without guarding as far as possible against the birth of handicapped children is indefensible. The clinical application of *in vitro* fertilization in all its forms demands research on embryos, but this should be undertaken only in full awareness of the complexities of parental consent. It is to be hoped that there will be a clarification of the legal status of these early embryos *in vitro* before long. □

1. Offences Against the Person Act, 1861, s.58.

2. Abortion Act, 1967, s.2.

3. Teifel, H.O. *New Engl. J. Med.* 294, 85-90 (1976).

4. Edwards, R.G. *Pontifical Academy of Sciences (Vatican)* 51, 193-249 (1984).

5. Congenital Disabilities (Civil Liability) Act, 1976, s.1.

*R.G. Edwards is at the Physiological Laboratories in Cambridge and M. Puxon is a practising leading Counsel.

Angle constraint for nuclear-pumped X-ray laser weapons

E. Walbridge*

Nuclear-pumped X-ray laser weapons are constrained to have an output beam divergence that is at least as great as a certain minimum value. Because this value is large, such weapons used for ballistic missile defence will require disturbingly high megatonnages in space: for example, a total of ≥ 73 megatons within range of 1,000 attacking missiles, with individual warheads up to 3.7 megatons or higher.

It has been suggested¹ that X-ray lasers, placed in space and pumped by the detonation of nuclear bombs, could be used to destroy Soviet missiles launched against the US. Pulses of radiation from the lasers would be directed at the missiles in their boost phase²⁻⁴. I now describe a fundamental constraint on the operation of such weapons: the divergence of the output beam must be as great or greater than a certain minimum value. This is a function of the wavelength, λ , of the output laser radiation and the length, L , of the laser rod. For $\lambda = 14 \text{ \AA}$ and $L = 2 \text{ m}$, the minimum value is 2.92×10^{-5} rad for the beam-divergence half-angle. Thus, at a target 2,000 km from the weapon, an output pulse will have spread out to cover a circular area of radius $\geq 60 \text{ m}$. Assuming a reasonable value for laser efficiency, I find that a sizeable nuclear bomb (such as ≥ 3.7 megatons (Mton) for a 50 laser-rod weapon) is required for laser pumping if the pulses are to be lethal against 50 fully hardened Soviet intercontinental ballistic missile boosters. At least 20 such weapons would be needed to defend against an attack of 1,000 missiles, for a total within-target-range warhead energy of > 73.2 Mton. With fewer boosters targeted by each weapon, the kilotonnage of each warhead can be reduced, but the number of weapons is increased (for a fixed number of attacking missiles) such that the total within-target-range warhead energy remains constant.

Unclassified literature data

X-ray laser weapons could be either based in space or lofted into orbit on warning of an attack¹. The latter approach now seems to be favoured³. According to Robinson¹, an X-ray laser weapon would consist of about 50 laser rods arranged in a 'ring' around a nuclear warhead. The weapon would operate as follows: (1) the pointing and tracking system would separately align each rod on a target; (2) once all the rods

had been aligned the warhead would be detonated; (3) X rays from the detonation would pump each of the 50 or so rods; (4) each rod would lase with the output pulses leaving the rods before they were totally destroyed by the expanding fireball; and (5) the laser pulses would propagate through space to their respective targets, destroying them by impulsive loading.

In a test of a nuclear-pumped X-ray laser by Livermore National Laboratory at the Nevada underground nuclear test site, a pulse of output laser radiation¹ was produced. This output pulse was of very high intensity ("several hundred terawatts"), of very short duration ("of the order of nanoseconds"), and consisted of X-ray radiation of 14 $\text{\AA}$ wavelength.

In the laser weapon, each rod would consist of "an atomically dense substance in solid form", and be essentially cylindrical in shape, with an estimated length of 3–8 feet and a "very small diameter"¹.

Apparently Livermore has tested a nuclear-pumped X-ray laser more than once, because Robinson⁵ later refers to "two partially successful underground tests". For X-ray laser weapons, critical issues include "demonstration of high efficiency and low beam divergence"⁶.

Generation of output pulse

The initial fireball created by the explosion of the warhead in the laser weapon can be taken to be a black body at 10^8 K . At that temperature, the photons emitted at the peak of the black-body curve have an energy of 43 keV, while the average photon energy is roughly 10 keV, which corresponds to a 1.24- $\text{\AA}$ wavelength. Thus the photons emitted by the fireball are in the X-ray range. In the fireball, electrons are coupled to ions by electrostatic forces, thus the two species will expand outward at essentially the same speed. Assuming a typical ion has the mass of ^{98}Sr (produced as a fission product), the velocity of outward expansion (v) to velocity of light (c) ratio will be $v/c = 0.0008$. Consider just one laser rod. It will be converted, by the intense flux of photons from the detonating warhead, to a plasma of electrons and multiply ionized ions. This

plasma will be transparent to X rays, in contrast to solid matter which is opaque. A population inversion will be created between certain ion energy states of low principal quantum number, n . Probably the laser will be a photoionization laser⁷⁻⁹, which achieves inversion by the preferential photoejection of electrons from inner atomic shells. Or it might be a plasma-recombination laser, which achieves inversion during recombination of a plasma¹⁰.

Three Russian authors¹¹ have considered photoionization lasers in which photons of $\sim 10 \text{ keV}$ or more pump ions of atomic number $Z \sim 30$ to inversions between $n = 3$ and 4, 5. The inversions occur because the $n = 3$ level is depopulated by photoejection more rapidly than the $n = 4$ and 5 levels. For $Z = 30$ (zinc), the $n = 5$ to $n = 3$ transition yields a photon of 14.2 $\text{\AA}$ wavelength¹¹, that is, essentially the output wavelength in the Livermore nuclear-pumped X-ray laser test¹.

Because of the low reflectivity of materials to normally incident X rays and the intense photon flux that would destroy mirrors, these cannot be used by a nuclear-pumped X-ray laser. Hence, the laser must achieve emission in a single pass.

The requisite single-pass amplification is produced by amplified spontaneous emission (ASE), ASE¹²⁻¹⁷ is just stimulated emission in a high-gain, single-pass, saturated amplifier with no feedback. The output pulse of an ASE laser consists of a superposition of component pulses — each arising from the amplification of a spontaneously emitted photon that happened to be directed approximately along the axis of the laser medium.

Beam-divergence angle

Before detonation of the warhead the laser rod is a cylinder of length L and diameter D , as shown in Fig. 1a. After detonation, the ion-electron plasma resulting from photoionization of the atoms in the rod can be taken to be in the form of a cylinder having length L and diameter D' (Fig. 1b). D' is constant with position, x , along the plasma cylinder and constant with time, t , between the moment

*Program in Science and Technology for International Security, Physics Department, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Present address: Argonne National Laboratory, Argonne, Illinois 60439, USA.

the laser pulse begins to amplify and the moment it leaves the rod. The diameter D' can differ from D .

A photon spontaneously emitted near the left end of the plasma cylinder in Fig. 1b and directed towards the right along the horizontal centreline GJ will be amplified, as will spontaneous photons similarly emitted on the left but directed towards the right along FK and HI. Thus the output pulse will be a superposition of elementary pulses, each of these being coherent and having resulted from the amplification of a distinct spontaneous photon. Neglecting diffraction, the angular divergence of the output beam will clearly have half angle D'/L rad.

The diffraction half angle, θ_{diff} , is given by $\theta_{\text{diff}} = 1.22\lambda/D'$. The total output beam half-angular divergence, θ_{tot} , consists of D'/L and θ_{diff} combined. There are two physically plausible ways to combine D'/L and θ_{diff} : they can be simply added together, or they can be combined according to the equation

$$\theta_{\text{tot}} = [(1.22\lambda/D')^2 + (D'/L)^2]^{1/2} \quad (1)$$

Equation (1) leads to a smaller value for the minimum beam-divergence angle than is obtained by just adding D'/L and θ_{diff} together, so we use equation (1) in the following. As $D' \rightarrow 0$ the diffraction term in equation (1) becomes arbitrarily large, while as $D' \rightarrow \infty$ the other term (the geometric divergence term) becomes arbitrarily large. Clearly, in between θ_{tot} has a minimum.

Setting the derivative of equation (1) equal to zero, one finds, for the minimizing value, D'_{mn} , of D' ,

$$D'_{\text{mn}} = (1.22\lambda L)^{1/2} \quad (2)$$

and for the minimum value, θ_{totmn} , of θ_{tot} , corresponding to D'_{mn} ,

$$\theta_{\text{totmn}} = (2.44\lambda/L)^{1/2} \quad (3)$$

For $\lambda = 14 \text{ \AA}$ and $L = 2 \text{ m}$, these two equations yield: $D'_{\text{mn}} = 58.4 \text{ }\mu\text{m}$, and $\theta_{\text{totmn}} = 4.13 \times 10^{-5} \text{ rad}$.

Now, equation (1) is based on the Fig. 1b model of the amplification process, in which model D' is taken to be constant as a function of x and t . In the general case, the diameter of the plasma column will be $D'(x, t)$, a function of both x and t . Will this dependence on x and t lead to a lower value of total output beam half-angular divergence than that given by equation (3)? I have considered all the different possibilities for the function $D'(x, t)$ and find that there is a subclass of all such functions for which total output beam half-angular divergence is less than the value given by equation (3). This subclass consists of functions: (1) of the form $D'(x, t) = D'(x)$ where $D'(x)$ is the diameter of a right cone as a function of the distance, x , along its axis (Fig. 2); and (2) for which $D'(L)$ lies within a certain range, R .

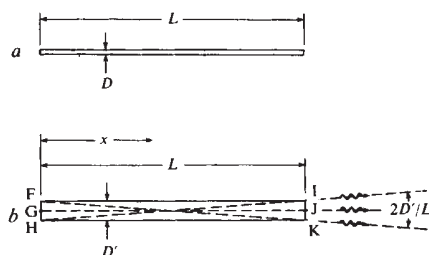


Fig. 1 a, The laser rod, of diameter D and length L , before detonation of the warhead. b, The plasma column, taken to be a cylinder of diameter D' and length L , into which the rod is converted through photoionization by photons from the fireball.

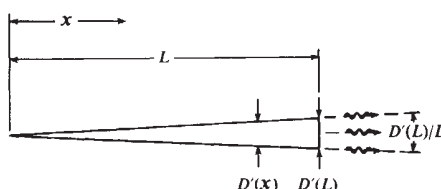


Fig. 2 A cone-shaped plasma column, shown in cross-section, yields the minimum beam-divergence angle.

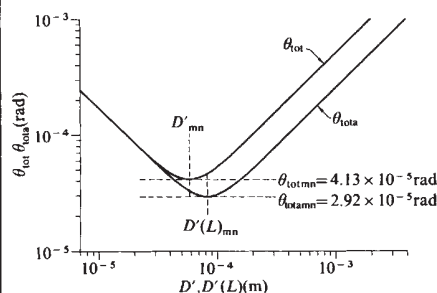


Fig. 3 θ_{tot} versus D' and θ_{tota} versus $D'(L)$. The half-angle beam divergence cannot be less than $\theta_{\text{totmn}} = 2.92 \times 10^{-5} \text{ rad}$ in the case of $\lambda = 14 \text{ \AA}$ and $L = 2 \text{ m}$.

The total half-angular beam divergence for a cone-shaped plasma column (Fig. 2), designated θ_{tota} , will be given by

$$\theta_{\text{tota}} = [(1.22\lambda/D'(L))^2 + (D'(L)/2L)^2]^{1/2} \quad (4)$$

For $D'(L) = D' = d$ (where D' refers to Fig. 1b), $\theta_{\text{tota}} \leq \theta_{\text{tot}}$ for all d . θ_{tota} and θ_{tot} are each plotted in Fig. 3 for $\lambda = 14 \text{ \AA}$ and $L = 2 \text{ m}$. In Fig. 3 the horizontal line corresponding to $\theta_{\text{totmn}} = 4.13 \times 10^{-5} \text{ rad}$ intersects the θ_{tota} curve at two points. For $\lambda = 14 \text{ \AA}$ and $L = 2 \text{ m}$, the range R consists of those values of $D'(L)$ lying between the abscissae of these two points, and for $D'(L)$ within this range $\theta_{\text{tota}} < \theta_{\text{totmn}}$. For general λ and L , R consists of those values of $D'(L)$ for

which $\theta_{\text{tota}}(D'(L)) < \theta_{\text{totmn}}$. The θ_{tota} curve attains its minimum value, θ_{totam} , at $D'(L)_{\text{mn}}$. Setting the derivative of equation (4) equal to zero, one finds,

$$D'(L)_{\text{mn}} = 2^{1/2} (1.22\lambda L)^{1/2}$$

and

$$\theta_{\text{totam}} = 2^{-1/2} (2.44\lambda/L)^{1/2} \quad (5)$$

For given λ and L , no function $D'(x, t)$ yields a value of total half-angular divergence smaller than θ_{totam} .

Note that θ_{totam} , given by equation (5), is the minimum total half-angle divergence one can attain with a nuclear-pumped X-ray laser. No matter what the actual, time-varying, shape of the plasma column, θ_{totam} is simply the best one can do in minimizing beam spread. Setting $\lambda = 14 \text{ \AA}$ and $L = 2 \text{ m}$ in equation (5) gives $\theta_{\text{totam}} = 2.92 \times 10^{-5} \text{ rad}$.

Rods at different targets

According to Robinson¹, "each lasing rod is separately aligned on a target by the pointing and tracking systems". Presumably this means that each rod is pointed at a different target. Assuming that to be the case I now explore the implications.

What is the maximum fluence, in J cm^{-2} , from a single laser rod and for given warhead energy, that would arrive at a target missile 2,000 km from an X-ray laser weapon? The maximum fluence is attained when the laser beam divergence half-angle is a minimum, that is, when it equals θ_{totam} . We make the following assumptions:

(1) $\lambda = 14 \text{ \AA}$; (2) $L = 2 \text{ m}$; (3) warhead energy = 10 kton = $4.18 \times 10^{13} \text{ J}$; (4) 70% of the warhead energy goes into X rays¹⁸; (5) there are N laser rods around the warhead and each rod/plasma column absorbs $1/N$ th of the total number of X-ray photons from the exploding warhead. (This assumption is conservative since only a fraction, < 1 , of the $1/N$ th photons will actually pass through the volume occupied by the rod/column, and not all of these will be absorbed. The validity of this assumption increases as the solid angle subtended by all N rods increases, that is, as N increases.); (6) each X-ray laser rod is 10% efficient, that is, the energy of the output X-ray pulse from each rod is 10% of the X-ray energy absorbed by the rod/plasma column.

The maximum fluence striking the target is then $(4.18 \times 10^{13} \text{ J}) (0.7) (0.1) / [N(2 \times 10^8 \text{ cm}^2)(2.92 \times 10^{-5} \text{ rad})^2 \pi] = (2.73 \times 10^4/N) \text{ J cm}^{-2}$. For $N = 50$, the value cited in ref. 1, $(2.73 \times 10^4/N) \text{ J cm}^{-2} = 0.55 \times 10^3 \text{ J cm}^{-2}$. Most of this energy will be absorbed by the target. I assume that it all is. The maximum fluence levels that present Soviet boosters can withstand without being destroyed, their hardness levels, lie in the range $0.4\text{--}2 \times 10^3 \text{ J cm}^{-2}$, according to ref. 6. However, a hardness of up to $20 \times 10^3 \text{ J cm}^{-2}$ could be achieved by future Soviet boosters⁶. Hence, the maximum fluence is

sufficient to 'kill' present Soviet boosters but not fully hardened ones.

An efficiency of 10% is quite optimistic. One would expect an X-ray wavelength photoionization laser to have a maximum efficiency of the order of 1% (M.A. Duguay, personal communication). If we take the target to be 2,000 km from the weapon, the efficiency to be 1%, with assumptions (1)–(5) above, unchanged, and $N=50$, the absorbed energy is only $0.055 \times 10^3 \text{ J cm}^{-2}$, a sub-lethal level for both present and hardened Soviet boosters. With a 1% efficiency and the target at 1,000 km, the fluence at the target will be $0.22 \times 10^3 \text{ J cm}^{-2}$, a still sub-lethal level for both. These estimates assume the minimum beam spread, $2.92 \times 10^{-5} \text{ rad}$, for $\lambda = 14 \text{ \AA}$ and $L = 2 \text{ m}$. In an actual nuclear-pumped X-ray laser weapon the spread would almost certainly be larger.

Hence, due to the intrinsic limit, θ_{total} , on the beam spread, a nuclear-pumped X-ray laser weapon with a 10-kiloton warhead and a laser efficiency of 1% will be ineffective in destroying hardened Soviet boosters, even if they are obligingly close. Of course, the warhead energy can simply be increased. Below the level at which the laser saturates this will yield an increase in the energy of the output pulse. For an efficiency of 1%, the target at 2,000 km, a 50-rod weapon, and assuming pulse energy varies linearly with warhead energy, a ≥ 73 -kton warhead would be required for lethality in the case of $0.4 \times 10^3 \text{ J cm}^{-2}$ hardness, and a ≥ 3.7 Mton warhead in the case of $20 \times 10^3 \text{ J cm}^{-2}$ hardness! The testing of a 73 kton weapon — equivalent to about 5 Hiroshimas — would at least be easy to monitor.

If a 3.7 Mton weapon were tested in space, how would it affect US and Soviet communication, control and surveillance satellites? And, via electromagnetic pulses, ground-based systems? These questions should be answered in the unclassified literature.

Rods at the same target

We now consider the case in which all N rods in an X-ray laser weapon are pointed at the same target.

In this case, a 10 kton warhead is sufficient to achieve lethality against present Soviet boosters, but not against fully hardened future ones, when assumptions (1), (2), (4) and (5) are satisfied, the laser rod efficiency is 1%, and the boosters are 2,000 km from the laser weapon. The same is true even when the target boosters are only 1,000 km away.

Scaling warhead energy up to achieve lethality, at 2,000 km range, against Soviet boosters fully hardened to $20 \times 10^3 \text{ J cm}^{-2}$, other assumptions as above, the requisite warhead energy is found to be ≥ 73 kton. Though considerably smaller than the 3.7 Mton found above for the fully hardened booster, multiply-targeted case, this energy is still sizeable. Further, with all 50 rods in

each weapon pointing at the same target, 50 times as many weapons are required as in the multiply-targeted case. The Soviets have 1,398 land-based intercontinental ballistic missiles¹⁹. If the boosters in these missiles are hardened to $20 \times 10^3 \text{ J cm}^{-2}$ and if 1,000 such hardened boosters were simultaneously launched against the US, at least 1,000 nuclear-pumped X-ray laser weapons, each having ≥ 73 kton warhead energy and a target booster within its 2,000 km range, would be required to kill the attacking boosters.

If the laser weapons are kept in orbit, as opposed to being 'popped up' from the Earth's surface on warning of attack, the total number of weapons will be much greater — at least by a factor of 25 — to always keep a sufficient number of weapons within range of the boosters as they move around their orbits. So, for orbital deployment at least 25,000 weapons are required, each of > 73 kton. This is for a weapon-to-booster range of 2,000 km. For a longer range, the number of orbiting weapons required will decrease but the kilotonnage of each must increase. What would be the environmental effects of placing 25,000 weapons into orbit? How would the release of 73×10^3 kton affect US space-based command, communication and control systems? And if the warheads on the 1,000 in-range weapons are not detonated nearly simultaneously, how will early detonations affect the pointing accuracy and overall functioning of the subsequently-to-be detonated warheads? How many 73-kton warheads would need to be detonated at one time in a test? One hundred, perhaps?

If, instead, the laser weapons are popped up, say from submarines, there may be insufficient time for the weapons to reach an altitude at which they can 'see' the attacking missiles, aim at them, and fire, before burnout of the missiles' boosters.

Groups of rods

Suppose the N rods in an X-ray laser weapon are divided into g groups of N/g rods each, all the rods in each group are aimed at the same booster, and different groups are aimed at different boosters. We have $1 \leq g \leq N$.

Let F_T denote the total fluence absorbed by a booster, at distance r from the laser weapon, due to all N/g rods directed at the booster. We have

$$F_T = 0.7E\epsilon/[g(r \times 2.92 \times 10^{-5})^2\pi] \quad (6)$$

where E is the warhead energy, and ϵ the laser rod efficiency.

Setting $F_T = 20 \times 10^3 \text{ J cm}^{-2}$, the maximum hardness of future Soviet boosters, substituting this value for F_T into equation (6), and solving for E when $\epsilon = 0.01$ and $r = 2,000$ km, yields

$$E = E_1 = g \times 73.2 \text{ kton} \quad (7)$$

where E_1 denotes the minimum warhead energy that would be lethal against a

maximally hardened future Soviet booster. For $N=50$ rods all pointing at different targets, $g=50$ and $E_1=3.7$ Mton, as noted above. For 50 rods all pointing at the same target, $g=1$ and $E_1=73.2$ kton.

If B boosters are simultaneously launched by the Soviets, the minimum number of within-range X-ray laser weapons, W , necessary to kill all the boosters, assuming no misses, is related to B by the equation.

$$Wg = B \quad (8)$$

Eliminating g from equations (7) and (8) yields

$$E_1 W = B \times 73.2 \text{ kton} \quad (9)$$

Therefore, the product of E_1 and W is proportional to B for given ϵ and r ; for $\epsilon=0.01$ and $r=2,000$ km, the constant of proportionality is 73.2 kton. Equation (9) states that the total within-range kilotonnage, $E_1 W$, on the W weapons, is independent of g . For example, if $B=1,000$ boosters, $\epsilon=0.01$, and $r=2,000$ km, then $E_1 W = 73.2$ Mton. No matter what the value of g , be it 50, or 1, or $1 < g < 50$, or $\leq N > 50$, there must be 73.2 Mton within range of the 1,000 boosters. This is a disturbingly large energy — whether placed in 20 weapons each having $E_1=3.7$ Mton ($g=50=N$), in 1,000 weapons each having $E_1=73$ kton ($g=1$), or in some intermediate number of weapons all with an intermediate $E_1=73.2$ Mton/ W ($1 < g < 50$). The 73.2 Mton in W weapons is required because of the large beam spread, which is at least 2.92×10^{-5} rad in half-angle.

However, this large beam spread has the advantage of reducing the need for precision pointing, since with that spread a lethal fluence would extend over an entire circular area of radius 60 m at 2,000 km.

I thank G. Marsh (IITRI), H. Haus (MIT), A. Javan (MIT), K. Tsipis (MIT), L. Allen (University of Sussex), and M. Duguay (Bell Laboratories) for helpful discussions and K. Tsipis for supporting this work.

1. Robinson, C.A. *Jr Aviat. Week* 114m 25-27 (1981).
2. Robinson, C.A. *Jr Aviat. Week* 119, 16-18 (1983).
3. Broad, W.J. *New York Times* 17 & 20 (15 November 1983).
4. Robinson, C.A. *Jr Aviat. Week* 119, 50-61 (1983).
5. Robinson, C.A. *Jr Aviat. Week* 119, 74-78 (1983).
6. Robinson, C.A. *Jr Aviat. Week* 119, 50-57 (1983).
7. Duguay, M.A. & Rentzepis, P.M. *Appl. Phys. Lett.* 10, 350-352 (1967).
8. Duguay, M.A. in *Laser Induced Fusion and X-Ray Laser Studies* (eds Jacobs, S.F. et al.) 557-579 (Addison-Wesley, Reading, (1976).
9. Silfvast, W.T., Macklin, J.J. & Wood, O.R. *Opt. Lett.* 8, 551-553 (1983).
10. Wood, O.R. II & Silfvast, W.T. *Appl. Phys. Lett.* 41, 121-123 (1982).
11. Bunkin, F.V., Derzhiev, V.I. & Yakovlenko, S.I. *Soviet J. Quant. Electron.* 11, 971-972 (1981).
12. Allen, L. & Peters, G.I. *J. Phys. A* 4, 238-243 (1971).
13. Allen, L. & Peters, G.I. *J. Phys. A* 4, 377-381 (1971).
14. Allen, L. & Peters, G.I. *J. Phys. A* 4, 564-573 (1971).
15. Allen, L. & Peters, G.I. *J. Phys. A* 5, 546-554 (1972).
16. Allen, L. & Peters, G.I. *J. Phys. Rev. A* 8, 2031-2047 (1973).
17. MacGillivray, J.C. & Feld, M.S. *Contemp. Phys.* 22, 229-310 (1981).
18. Glasstone, S. (ed.) *The Effects of Nuclear Weapons*, Rev. edn (US Atomic Energy Commission, Washington DC, 1962).
19. Forsberg, R. *Scient. Am.* 247, 52-61 (1982).

Immunology made accessible

In immunology, abstract concepts are being made tangible by molecular structures. One new development shows that protein structure can evolve without changing common architecture.

THERE are two ways of regarding the recent spate of understanding gathered by molecular biologists of the molecules characteristically involved in the immune response of vertebrates: either more data mean that a field which is already too confusing may be further confused, or the point may have been reached at which confusion is dispelled. What follows is based on the assumption, hopeful though it may be, that the second is the better approximation to the truth, and that two articles elsewhere in this issue (pp.233 and 235) bear this out.

As in many other circumstances, concepts that at first must perforce be described in abstract terms (*gene* for example, or for that matter *electron*) are not easily turned over in the mind, or argued about. To begin with, the definitions of crucial concepts will be, at best, partial definitions. Historically, the electron began life as a carrier of electricity, charge and mass unspecified, and became a tangible entity only when J.J. Thomson had shown that the ratio of charge to mass is a constant. Genes began as abstractions, the propensity of an organism to have an external attribute of some kind, with only a few rules restricting the mechanism of their inheritance.

The classical concepts of immunology, *antigen* and *antibody*, have a similar and in many ways even more confusing history. Antigens began life empirically as materials that would evoke a neutralizing immune response in vertebrates into which they are injected, and antibodies were the agents of that neutralizing activity. This circular definition nevertheless had the merit of explaining a variety of phenomena, if predicting few.

Throughout the classical period in immunology, the outstanding puzzle was the capacity of the vertebrate immune system to respond specifically to an apparently infinite variety of antigens, some of them made only recently by organic synthesis and which thus can have played no part in animal evolution. Only in the past decade has this process been made tangible by the recognition that antibody molecules have a common molecular architecture, but that the common plan includes regions where the amino acid sequence may be widely varied as a consequence of the ways in which bits and pieces of genes are physically rearranged within the lymphocytes called B cells.

So here is one component of the immune

system made tangible. Immunoglobulin molecules are Y-shaped structures consisting of two identical halves, with the vertical stem of the Y anchored in the external membrane of the cell producing it. The anchor has the same amino acid structure for each class of immunoglobins. It is generally assumed that the immune system makes all possible combinations of the bits and pieces which constitute the variable genes, but that those which appear in measurable amounts are those required by the exigencies of survival. Little is known for certain of the degree to which the diversity provided by gene rearrangement is supplemented by mutation *in situ*, or of the mechanism by which genes are rearranged.

This is the stuff of which primers are written. Most of the excitement in the past four years that immunologists have mostly shared among themselves has centred on the lymphocytes called T cells, which produce no antibodies but which either kill off cells carrying antigens characteristic of, say, a virus infection or which assist the immune response in other ways.

So if T cells must have ways of recognizing antigens, what more natural than that they too should be constructed from the genes that make antibodies? Not so. By two years ago (see Jensenius, J.C. and Williams, A.F. *Nature* 300, 583; 1982), all that was known of the "T-cell receptor" was that no substantial part of it has much in common with the immunoglobins.

What had however become apparent was that T cells will recognize antigens on other cells only if these are associated in the cell membrane with the antigens which control the histocompatibility (skin grafting) reaction; presumably the histocompatibility antigens which the body has learned to tolerate seem like those belonging to some other individual when associated with, say, a protein produced by a virus infection. And people were already surmising (Robertson, M. *Nature* 297, 629; 1982) that T-cell receptors and the histocompatibility antigens are anchored in the cell surface in much the same way as antibody molecules, although the sources of variability and thus specificity are, of course, not gene rearrangements.

Earlier this year, the common features of the architecture of all these molecules — immunoglobulins, T-cell receptors and the two classes of histocompatibility antigens — were reasonably well understood. The class I antigens are, for practical purposes, those that provoke a cytotoxic immune

response, the class II antigens those that stimulate the helper function. Nothing, however, is known of the ways in which these antigens interact with foreign antigens in the membrane of, say, virally infected cells to produce their immunological provocation. At the end of June, however, the structure of what seems to be the T-cell receptor was described (Saito, H. *et al. Nature* 309, 757; 1984).

The article by P. Travers *et al.* (this issue, p. 295) is something of a diversion, but a cheerful one. What this group has done is to use a computer model to infer the structure of membrane-bound molecules of class II histocompatibility from their known (or inferred) amino acid sequence. Put simply, the conclusion is arresting: although the detailed correspondence between the structure of these molecules and those of the immunoglobins is insubstantial ("low homology" is what the molecular biologists would say), the general architecture of the molecules is very similar. Again, there is a stem anchored in the cell membrane by the hydrophobic character of its amino acids. Again, the external region is that which carries the variable (and thus the specific) part.

In spite of the element of circularity in this argument stemming from the way in which the structure of antibody molecules has been allowed to guide the model-building, the significance of this neat construction should not be overlooked. In passing, it says much about the power of computers as tools for building molecular models. More pointedly, it is a powerful demonstration that two molecules may differ considerably from each other in their detailed structure and yet have their general architectural shape in common.

Inevitably, as the authors point out, this bears directly on the evolution of the different components of the immune system. The rules of the game are that the changes may be rung on successive nucleotides in the genes, and thus on the amino acids, subject only to the constraint that the architecture of the molecules should be preserved. Their estimate that it has taken 500 million years to accomplish the specialism and diversity of the vertebrate immune system is less striking than their repetition of the suggestion that the immune system has evolved from some earlier means by which primitive aggregates of cells, sponges or corals, distinguished members of their own species.

John Maddox

Botany

Manipulation of photosynthesis

from J. Barber

BEFORE the goal of creating improved cultivars of crop plants by *in vitro* genetic engineering can be reached, it will be necessary to establish appropriate molecular biological techniques for precise gene transfer and to identify the key physiological processes and biochemical reactions that must be changed to produce the new varieties. In response to these challenges Britain's Agricultural and Food Research Council (AFRC) sponsor two priority programmes. One is concerned with developing techniques for the genetic manipulation of plant cells (see *Nature* 304, 498; 1984). The groups that contribute to the other programme, which aims to elucidate the processes that limit photosynthesis and plant growth, met recently to consider the outcome of its first five years. A clear message from the meeting* was that there has been a substantial increase in knowledge which is now starting to form a firm basis for closer interaction with the molecular geneticists.

Our knowledge of the organization and functioning of the light-harvesting system and associated electron-flow processes has advanced to such a level that it is now possible to give rational explanations to the mechanisms by which some, but not all, plants can adjust to growth at different temperatures and under different lighting conditions. Both of these environmental factors are commercially important to the agricultural and horticultural industries. As emphasized by C. Pollock (Welsh Plant Breeding Station, Aberystwyth), the control of photosynthesis and plant growth is not just at the level of electron transport but also involves optimization of carbon metabolism and the adoption of various strategies for partitioning assimilates.

It is seldom clear which specific proteins should be the (indirect) targets of genetic manipulation in order to create new crop plants but there are already some examples where molecular biology is having considerable impact. For example, there is the 32,000-molecular weight protein which is a structural component of the photosystem II complex and is known to bind a wide range of commercially important herbicides, including the triazines, ureas and nitrophenols. This hydrophobic protein is coded for in the chloroplast genome and its amino acid sequence has been determined from the gene sequence (Zurawski, G. *et al.* *Proc. natn. Acad. Sci. U.S.A.* 79, 7699; 1982). J. Hirschberg and L. McIntosh have sequenced the same gene, but isolated from the atrazine-resistant biotype of the weed

Amaranthus hybridus, and conclude that the resistance can be accounted for by a single amino acid change (*Science* 222, 1346; 1983). According to Alison Telfer (Imperial College, London) there is little difference in the photosynthetic capacity of atrazine-resistant and atrazine-susceptible plants of four different species, suggesting it would not be detrimental to 'engineer' atrazine-resistance into crop plants.

Techniques for specifically manipulating the chloroplast genome have not so far been developed, but site-directed mutagenesis is proving to be a valuable approach when altering plant proteins synthesized by gene expression in bacteria, in particular the enzyme ribulose biphosphate (RUBP) carboxylase. It is this enzyme which fixes CO₂ by using the reducing power and ATP generated by light-induced electron transport. In higher plants the enzyme is a complex consisting of eight large and eight small subunits while in some photosynthetic bacteria the enzyme structure is simpler. As G.H. Lorimer (Du Pont de Nemours Co., Wilmington) outlined in an overview lecture, the carboxylase activity of this enzyme requires the initial formation of a chemical complex of CO₂ and a divalent cation. Its activation and subsequent catalysis are affected by a range of metabolites, the significance of which is not yet fully understood for the *in vivo* system. Perhaps the most striking and important feature of the enzyme is its ability to catalyse an alternative reaction involving the oxygenation of sugar biphosphate. Both the carboxylase and oxygenase reactions occur at the same site and compete with each other. The oxygenase reaction produces phosphoglycolate and the metabolism of the resulting hydroxy acid leads to release of CO₂ in a process known as photorespiration.

Because this alternative pathway is energetically wasteful, especially at limiting light intensities, it is a goal of the AFRC programme to manipulate the enzyme genetically so as to increase its carboxylase activity and decrease its oxygenase activity. Site-directed mutagenesis is the obvious approach, and forms the basis of some encouraging work by S. Gutteridge (Rothamsted Experimental Station) in collaboration with G. Lorimer, on the *in vivo* site-directed mutagenesis of the RUBP carboxylase of the photosynthetic bacterium *Rhodospirillum rubrum*. Although this enzyme consists of only two identical subunits, its amino acid sequence is remarkably similar to that of RUBP carboxylase from higher plants, particularly in a region of acidic amino acids involved in activation of the enzyme. At residue 198 there is a highly conserved

aspartic acid which Gutteridge and Lorimer chose to replace by glutamic acid. The modified enzyme, synthesized in *Escherichia coli*, was found to have lower carboxylase and oxygenase activities as well as a modification in its divalent cation-binding properties. This result not only provides further evidence for the common catalytic site of carboxylase-oxygenase activity but also represents the starting point for a series of experiments to manipulate the enzyme and elucidate the molecular mechanism by which it functions.

Precisely this approach is also needed to study the RUBP carboxylase-oxygenase activities of higher plants, although a complication arises because the small subunits of the enzyme are encoded in the nuclear DNA while the gene for the large subunit is located in the chloroplast where final assembly takes place. Even so, A. Gatenby (Plant Breeding Institute, Cambridge) could report that genes for the small and large subunits have been cloned and separately expressed in *E. coli*. Unfortunately the products did not assemble into an active enzyme. It may be necessary to introduce both genes into the same *E. coli* cell to overcome this problem. One approach would be to isolate appropriate genes from cyanobacteria which have RUBP carboxylase similar to higher plants but do not have the complication of differentiation into chloroplast and nuclear DNA. In any event it seems highly likely that *in vitro* methods will soon produce active RUBP carboxylase so that the powerful technique of site-directed mutagenesis can be exploited. To aid these developments detailed structural studies of the crystallized enzyme by X-ray diffraction will be useful and progress in this direction was reported by Professor Sir David Phillips' group from Oxford.

I left the meeting feeling optimistic that we would eventually be able simultaneously to reduce oxygenase activity of RUBP carboxylase and to increase the efficiency of the carboxylation step. It remains to be seen, however, whether the overall efficiency of photosynthesis in the intact plant will thereby be increased, as hoped, or whether photorespiration turns out to be obligatory for effective carbon assimilation by the C3 pathway. □

J. Barber is Professor of Plant Physiology and Head of the ARC Photosynthesis Research Group at Imperial College, London SW7 2BB.

100 years ago

THE first mail from Kadiak Island received this season has arrived at San Francisco, bringing dates to May 2. According to the correspondent of the *Bulletin*, the account of the eruption of the volcano on Augustine Island, Cook's Inlet, sent by the last advices of 1883, was much exaggerated. The island "was not split in two, and no new island was formed but the west side of the summit has fallen in, forming a new crater, while the whole island has arisen to such an extent as to fill up the only bay or boat harbour." No tidal waves were observed on the west shore of Cooke's Inlet or on Kadiak Island.

From *Nature* 30, 275, 17 July 1884.

*Agricultural and Food Research Council's Meeting on Photosynthesis, 15-17 April, University of Sheffield, where Professor David Walker has recently established a Research Institute of Photosynthesis.

Meteorology

Stratospheric planetary waves

from David Andrews

FOR several decades, meteorologists have known that the winter stratosphere is frequently disturbed by wave-like events of planetary scale. Recently, great progress has been made in exploring the dynamics of these waves and understanding their influence on the stratosphere as a whole. These advances have been brought about by a combination of observational studies based on data from satellites, 'experiments' with computer models of the stratosphere and theoretical investigations.

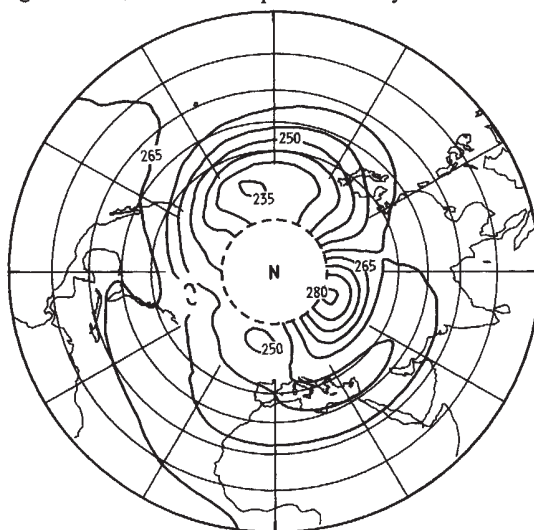
Much of the impetus for the exciting developments in this area has come from the increasing availability of large amounts of data — with worldwide coverage over several years — provided by orbiting satellite instruments. We now possess extensive data-bases, including measurements of stratospheric temperatures, winds, pressures and the concentrations of chemical constituents such as ozone and methane. The figure shows a stereographic map of a typical stratospheric planetary wave, as revealed by temperature measurements from Oxford University's Stratospheric and Mesospheric Sounder on the Nimbus 7 satellite.

In parallel with these observational studies, there has been an increased use of numerical models of the stratosphere. These models solve the mathematical equations which (to a good approximation) express our current understanding of the basic physical principles governing atmospheric behaviour. One use of models is to attempt to predict the future development of the stratosphere, rather as weather-forecasting models are used for the troposphere. Models can also perform idealized hypothesis-testing experiments, to refine our knowledge of the dominant mechanisms controlling the structure of the stratosphere.

The large amounts of data provided by these observational and computational studies must be organized into forms suitable for analysis by stratospheric meteorologists. A popular approach is to separate each atmospheric field into a 'zonal mean' part (obtained by averaging around latitude circles) and a 'wave' part, given by the departure from that mean; another method uses transient deviations from a time mean. The most useful average will depend on the phenomenon under consideration, and in some cases no simple average may be particularly appropriate.

The basic linear theory of planetary (or Rossby) waves propagating on a zonally-symmetric mean flow has been well estab-

lished for many years. Only comparatively recently, however, has the two-way non-linear interaction between these waves and the mean flow received much attention. In common with other examples of wave-mean flow interactions in fluids, it is a subject in which naive physical arguments — perhaps based on inappropriate analogies with simpler wave systems — can sometimes be seriously misleading¹. Physical interpretation must be supported by a quantitative theoretical framework, and current developments in dynamical



Polar stereographic map of stratospheric temperature (degrees Kelvin, contour interval 5 degrees) at 1 mbar (near 45 km altitude) on 7 January 1982. Inner circle, 70°N; outer circle, Equator. (Courtesy of M. Corney, Oxford.)

meteorology are going some way towards providing such a framework.

Several recent studies illustrate the variety of ways in which meteorologists are handling and interpreting stratospheric and tropospheric data. In a paper² in *Journal of the Atmospheric Sciences*, Hartmann, Mechoso and Yamazaki use the 'zonal mean/wave' separation, for which the theory of wave-mean flow interaction is now quite well developed³. Previously, authors have contrasted the quasi-stationary planetary waves dominating the northern stratospheric winter with the eastward-travelling disturbances which are the most prominent waves in the southern stratospheric winter⁴. The novel feature of Hartmann *et al.*'s work is that it focuses on data from the Southern Hemisphere; it is the first detailed quantitative analysis of Southern Hemisphere waves and their interaction with the zonal-mean flow.

Another pair of papers, by McIntyre and Palmer^{5,6}, use no explicit averaging, but present satellite-derived maps of 'Ertel's potential vorticity' on a roughly horizontal

surface of constant entropy in the mid-stratosphere. Both potential vorticity and entropy are approximately conserved by air parcels as they move, and the maps provide blurred but fascinating pictures of stratospheric motions during a large-amplitude planetary-wave event in January 1979. They suggest that large air masses are rapidly and irreversibly deformed during this event into long thin filaments, which may in turn break up into eddies. McIntyre and Palmer call this process 'planetary-wave breaking', by analogy with the breaking of ocean waves on a beach; they also propose that the theory of 'nonlinear critical layers'^{7,8} may model some aspects of the dynamics of such events. More work still needs to be carried out to confirm the reliability of potential vorticity maps, to assess the significance of wave-breaking events for the stratosphere as a whole and to understand their dynamics in detail. Nevertheless, the papers of McIntyre and Palmer open up interesting new possibilities for studying stratospheric waves.

Similar developments are taking place in research into tropospheric waves. For example, Hoskins, James and White have developed a theory for interpreting the interaction of transient cyclones with the time-averaged flow⁹. This approach may prove useful in the analysis of 'blocking' situations, such as that responsible for the British drought of 1976; isentropic potential vorticity maps are beginning to be of diagnostic value here too.

These recent studies highlight the use being made by meteorologists of an increasing number of theoretical tools for probing the behaviour of atmospheric waves, thereby enlarging our knowledge of the principles of large-scale atmospheric dynamics. A better grasp of fundamentals will in turn be of enormous aid in the solution of such vital practical problems as modelling the transport of chemical pollutants which might threaten the ozone layer and predicting the influence of increased atmospheric carbon dioxide on climate. Even weather-forecasting will surely benefit from a deeper understanding of the dynamics of the atmosphere. □

1. McIntyre, M.E. *J. Fluid Mech.* **106**, 331 (1981).
2. Hartmann, D.L., Mechoso, C.R. & Yamazaki, K. *J. Atmos. Sci.* **41**, 351 (1984).
3. Edmon, H.J., Hoskins, B.J. & McIntyre, M.E. *J. Atmos. Sci.* **37**, 2600 (1980).
4. Leovy, C.B. & Webster, P.J. *J. Atmos. Sci.* **33**, 1624 (1976).
5. McIntyre, M.E. & Palmer, T.N. *Nature* **305**, 593 (1983).
6. McIntyre, M.E. & Palmer, T.N. *J. Atmos. Terr. Phys.* (in the press).
7. Stewartson, K. *Geophys. Astrophys. Fluid Dyn.* **9**, 185 (1978).
8. Warn, T. & Warn, H. *Stud. Appl. Math.* **59**, 37 (1978).
9. Hoskins, B.J., James, I.N. & White, G.H. *J. Atmos. Sci.* **40**, 1595 (1983).

David Andrews is in the Department of Atmospheric Physics, University of Oxford, Clarendon Laboratory, Parks Road, Oxford OX1 3PU.

Astronomy

Quasar polarization properties

from Robert Carswell

Two surveys of quasars recently undertaken by astronomers at Steward Observatory in Arizona and reported in the *Astrophysical Journal* (279, 465 and 485; 1984) give us a comprehensive overview of the polarization properties of the quasar continuum radiation at optical wavelengths. Some of the results are valuable confirmations of old beliefs, such as the strong link between high polarization and rapid variability at optical and radio frequencies, but the work is notable for the new material it provides. One of the most important conclusions to emerge is that quasars can be divided into two classes, those with linear polarization of less than 2 per cent and a small fraction (about 1 per cent of the total number) with polarization greater than 3 per cent and up to an observed maximum of 30 per cent. In the case of the weakly polarized sources, the polarization appears to be aligned with the extended radio structure, where it is present, and there is evidence that the polarization is greater at short wavelengths than at long ones.

One of the popular mechanisms for producing continuum radiation in quasars and Seyfert galaxies (which behave very much like mini-quasars) is radiation from relativistic electrons which are constrained by a magnetic field (synchrotron radiation). If the magnetic field is ordered, the radiation emitted by this process should be highly polarized — up to about 70 per cent — and the degree of polarization should not depend strongly on wavelength. Astronomers seeking to confirm the importance of synchrotron radiation at optical wavelengths were rewarded with a small number of quasars which show strong rapidly-variable polarization, but it soon became clear that in most cases the polarization was below the few per cent detection limit of the early studies. Moreover, while synchrotron radiation at optical wavelengths adequately explains those quasars which show high polarization, for the large majority it is not clear that this mechanism is operating, or even that the optical continuum is necessarily non-thermal in origin.

The Steward Observatory polarization studies go some way towards answering some of the outstanding questions. First, they show that linear polarization is a feature of the continuum radiation at optical wavelengths of all quasars, albeit at a rather lower level than might have been expected. Second, on the basis of the new results, several proposed explanations for the polarization properties of quasars can now be discounted. It has been suggested that all quasars have the same basic structure, but that the properties observed are

determined by our orientation with respect to some anisotropic component. However, we would not then expect to see a clear differentiation in polarization properties between the bulk of quasars and the few highly polarized ones. Nor does this observation support a model in which a polarized (synchrotron?) component interacts with some other (thermal?) continuum, with the relative strengths varying from quasar to quasar. Rather, the two classes of polarization appear to reflect differences in the intrinsic properties of the quasars.

The observations on the low-polarization quasars are, in fact, in reasonable agreement with the predictions of only one model — polarization by dust. (A similar model is used to describe our own Galaxy.) Light scattered by dust grains becomes polarized, independent of the alignment of the grains, but dependent on the angle through which the light is scattered. If it were not possible to resolve the dust-scattering region from the continuum source, as would be the case for quasars, the object itself would appear polarized, provided the dust did not have a spherically symmetric distribution. Thus, if the system is rotating and the dust is in a disc or flattened sphere, for example, we should see a

net polarization except in the relatively few cases in which the axis of rotation is aligned towards us. Moreover, since dust grains scatter light more effectively at short wavelengths than at long wavelengths, short-wavelength light is more polarized. This is exactly what Stockman and his colleagues at Steward Observatory have seen. They also find some alignment of the radio structure with the polarization, assuming that the rotation axis of the dusty disc can be aligned with the radio structure.

This is a very important result, and serves to strengthen the link between quasars and Seyfert galaxies, whose emission properties have been explained in terms of dust obscuration. As far as understanding the nature of the continuum radiation, however, it is perhaps the least helpful of the explanations which have been considered. Since the dust is almost certainly outside the continuum source and will modify any intrinsic polarization properties the continuum might have, the polarization observations provide few constraints on a model of the continuum radiation. Even synchrotron radiation from regions with a very disordered magnetic field remains a possibility. Future studies of the highly-polarized and rapidly-variable quasars will no doubt yield further valuable information for that class, but it is not clear how they relate to the low-polarization objects. □

Robert Carswell is at the Institute of Astronomy, The Observatory, Madingley Road, Cambridge CB3 0HA.

Drug delivery

Old drugs in new clothing

from Gregory Gregoriadis

WITH the cost of the development of new drugs forever mounting, it may be more sensible to make better use of those already available. To this end, the controlled delivery of current drugs by biodegradable carriers could decrease their side effects, dosage and frequency of administration. A recent meeting* was devoted to the consideration of alternative carrier systems for drugs, including man-made materials, such as polymers and liposomes, and natural materials, notably antibodies. They can be used in the form of implants, releasing drugs at rates conducive to optimal action, often through controlled degradation, or be employed in systems of targeted delivery.

There has been no shortage of ingenuity in controlling degradation of implants. J. Heller (SRI, Menlo Park), for instance, uses poly(ortho)ester polymers with acid-sensitive linkages in their backbone, incorporating acidic salts in a highly hydrophobic environment. Since only salt on the sur-

face of the polymer is exposed to water, polymer erosion by hydrolysis should occur only at the surface. Therefore, drug release can be controlled over periods ranging from hours to months by the amount of salt present. Controlled release can also be achieved by a combination of drug diffusion and polymer biodegradation. Thus, microspheres made of copolymers of lactic acid and glycolic acid will release a drug as a function both of its concentration and of certain physical parameters of the polymer; injected microspheres loaded with a contraceptive steroid led to predictable and efficacious levels of the steroid in volunteers for up to six months (D.H. Lewis, Stolle Research and Development Corporation, Alabama). Even non-biodegradable materials, such as polymethyl methacrylate or polystyrene, can be converted into useable carriers by introducing biodegradable sections along the polymer chain (W.J. Bailey, University of Maryland).

An alternative to the passive entrapment of drugs in a polymer matrix is to arrange for them to be bound on to the matrix and

* 'Macromolecules as drugs and as carriers for biologically active materials', New York, 26–28 March, was organized by Drs D.A. Tirrell and L.G. Donorum and sponsored by the New York Academy of Sciences.

released through hydrolysis. Indeed, polymers can be made to consist entirely of the drug, herbicides for example. Control of hydrolysis will depend on such variables as the hydrophilicity of the polymer (F.W. Harris, University of Akron), the nature of the bond and its distance from the polymer's backbone (C.L. McCormick, University of Southern Mississippi). In a related approach, J. Kopeček (Institute of Macromolecular Chemistry, Prague) and J.B. Lloyd (University of Keele, UK) have attached drugs onto hydroxypropyl methacrylamide via oligopeptide sequences that are stable in the circulating blood but degraded within the lysosomes of the target tissue.

In contrast to the implant version of drug delivery, targeted systems are mobile; moreover, they can be designed so that upon binding to the target tissue, local factors release the drug. Killing of malignant and other unwanted cells is among the tasks considered feasible by this approach. A much explored version of targeted delivery uses tissue-specific antibodies to carry cellular toxins to their targets. Less explored are such systems as that in which the selective uptake of a drug-carrier by parenchymal cells of the liver *in vivo* is achieved by attaching to the carrier a ligand that terminates in galactose and so becomes bound to the galactose receptors of the cells (G. Gregoriadis, University of London.) Drug delivery to these cells could be of value in a number of diseases including hepatitis B and the liver stage of malaria.

Several groups have succeeded in targeting liposomes via antibodies. In fact, as antibody-bearing liposomes have a higher valency than the soluble ligand and can bind to cells with a greater affinity constant (D. Papahadjopoulos, University of San Francisco), their use for the specific killing of cells may be preferable to that of free antibodies. Furthermore, because liposomes will accommodate almost any available drug, a variety of cytotoxic mechanisms can be designed to suit particular needs. For example, H. Bayley (Columbia University) reported the selective killing of human T lymphocytes on exposure to UV light when in the presence of liposomes that are coated with an antibody against T lymphocytes and contain a phototoxic agent. In a rather intriguing variation of targeting, reported by R.L. Juliano (University of Texas Medical School), the toxicity of amphotericin B was reduced by its incorporation into liposomes without any loss of its antifungal activity. Apparently, liposomal amphotericin transfers to the fungal cell walls and membranes but not to mammalian cells.

Tailoring of liposome composition (for example, by a combination of excess cholesterol and high-melting phospholipids) can help to stabilize them in the presence of blood. A new approach to stability is based on the use of lipids containing diacetylene, butadiene or styrene groups. When incorporated into liposomes

and exposed to UV light, the lipids form polymer chains through cross-linking (D. Chapman, University of London). The resultant 'polymerized' liposomes are exceedingly stable and retain entrapped solutes in the presence of organic solvents and detergents (S.L. Regen, Marquette University, Milwaukee; D.F. O'Brien, Eastman Kodak Co., Rochester), though conditions can be arranged for drug release to occur *in situ* (H. Ringsdorf, University of Mainz). Paradoxically, polymerized liposomes may prove too stable and, hence, toxic for use *in vivo*.

The meeting made clear how much polymer chemists and biologists need each other in their endeavour to control drug action: more often than not, chemists will design a delivery system without proper reference to the biological milieu for which the system is intended, and medical researchers will be unable to conceive or implement structural alterations in the system conducive to optimal function. □

Gregory Gregoriadis is in the Academic Department of Medicine, Royal Free Hospital School of Medicine, London NW3 2QC.

Earth science

New date for diamonds

from Nick Rogers and Chris Hawkesworth

A PAPER in this issue of *Nature*¹ dating inclusions in diamonds from southern African kimberlites to more than 3,000 Myr old is important for two reasons. Scientifically, it is of interest because it provides significant insight into the origins of the diamonds themselves and the nature of the Archaean mantle. It is also of commercial interest because it has implications for successful prospecting.

Diamonds are an occasional component of very rare igneous host rocks called lamproites and kimberlites. Two explanations have been advanced for their origins: either diamonds crystallized from the magmas which cooled to form the igneous rocks in which they are now found; or, those magmas picked up diamonds as exotic fragments, accidentally entrained during violent volcanic activity. The former suggests that successful prospecting is simply a question of finding lamproites and kimberlites, whereas the latter implies that a more sophisticated approach is needed. Advocates of the 'exotic fragments' hypothesis have been better rewarded.

Most kimberlites and lamproites are less than 200 Myr old but, intriguingly, the majority intrude rocks as old as 3,500 Myr. They are not all diamondiferous; those that are appear to be confined to the older cratonic areas of continents², which suggests that diamonds too may be billions of years old. Moreover, the distribution of diamonds is not simply related to the distribution of particular igneous rocks, which is at least circumstantial evidence that diamonds and their host rocks are not genetically related.

Diamonds have never been dated satisfactorily. Nevertheless, some of them contain minute inclusions of silicate minerals, commonly olivine, pyroxene and garnet, similar to the minerals found in the upper mantle. These minerals in turn contain very small quantities (< 10 µg g⁻¹) of trace elements involved in radioactive decay schemes, and it is these that have now been analysed by Richardson *et al.*¹. ¹⁴⁷Sm decays to ¹⁴³Nd and the present-day isotope

composition of neodymium in the inclusions gives a date of 3,200–3,400 Myr old.

The immediate importance of the new study is that it provides, for the first time, convincing evidence that segments of Archaean mantle survived at least until the time of kimberlite formation (90 Myr) and presumably until the present day. The diamond hosts act as protective shields around the inclusions, so no age information is lost by interaction between the inclusions and, for example, the kimberlite magma. Moreover, because the age is that of the silicate inclusions, it is independent of speculation as to whether the host diamonds are cogenetic. However, if the diamonds too are 3,000 Myr old, then not only will these results influence future prospecting strategy, but they may also imply that at least beneath Kimberley, the continental lithosphere has been over 150 km thick for 3 × 10⁹ years. [Old segments of the upper mantle are thought to occur within the lithosphere, because it is relatively cold and rigid and so can age undisturbed by convection, and for likely geothermal gradients diamond is only stable below 150 km (ref. 3).] In which case, models for the present-day distribution of heat flow in the continental lithosphere would also be relevant to the Archaean⁴.

Cathodoluminescence and carbon-isotope studies have shown that diamonds are complex minerals. Individual crystals exhibit concentric zonations and carbon-isotope (δ¹³C) variations between core and rim of up to 5‰ (refs 5–8). This variation suggests that diamonds may experience more than one period of growth, and is even consistent with the possibility that small diamonds, and the rims of larger stones, crystallized from host kimberlite.

The physical relationship between the diamonds and their inclusions gives some insight into their relative ages. Usually, it is only possible to conclude that inclusions are older than their host minerals, with little indication of the time interval involved. Sometimes, however, the crystal form of silicate inclusions seems to resemble more

closely the internal structure of diamond than that of silicate minerals. How common that is, or whether it is most often seen in inclusions that have a similar crystallography to diamond, such as garnet, is unknown, but when present it is regarded as compelling evidence that the diamonds and inclusions are truly cogenetic. Further studies are necessary to clarify the relationship between inclusion morphology and the structural complexities of diamond, but it seems likely that both the inclusions and the diamonds worked on by Richardson *et al.*¹ are 3,200–3,400 Myr old.

These ages are consistent with the observation that diamond-containing kimberlites only occur in old continental areas. The implication is that most diamonds are old, which is perhaps surprising and certainly prompts speculation as to why that might be. One suggestion is that diamonds are confined to old areas because it is only beneath those portions of the continental crust that the lithosphere is thick enough to extend to the depths where diamonds are stable². But why should the lithosphere that stabilized in the Archaean be thicker than that formed more recently?

An alternative approach is to seek a

chemical explanation. The carbon content of the mantle was presumably greater early in Earth history than at present. Thus, enrichment processes that scavenge and concentrate trace elements in the upper mantle¹⁰ would have generated much higher localized concentrations of carbon in the Archaean than in recent times. We believe that the distribution of natural diamond², and its extremely old ages¹, is more likely to be related to the high carbon content in the Archaean mantle than to differences in lithosphere thickness. Some diamonds appear to have been derived from recycled 'crustal' carbon⁶, and we speculate that these are the ones most likely to yield younger (post-Archaean) ages. □

1. Richardson, S.H. *et al.* *Nature* **310**, 198 (1984).
2. Gurney, J.J. & Harte, B. *Phil. Trans. R. Soc.* **A297**, 273 (1980).
3. Boyd, F.R. & Nixon, P.H. *Phys. Chem. Earth* **9**, 431 (1975).
4. Davies, G.F. *Earth planet. Sci. Lett.* **44**, 231 (1979).
5. Deines, P. *Geochim. cosmochim. Acta* **44**, 943 (1980).
6. Milledge, H.J. *et al.* *Nature* **303**, 791 (1983).
7. Swart, P.K. *et al.* *Nature* **303**, 793 (1983).
8. Matthey, D. *et al.* *Nature* (submitted).
9. Boyd, F.R., Gurney, J.J. & Nixon, P.H. *EOS* **64**, 341 (1983).
10. Hawkesworth, C.J. *et al.* *Nature* (in the press).

Nick Rogers and Chris Hawkesworth are in the Department of Earth Sciences, The Open University, Milton Keynes MK7 6AA.

Ecology

Rise and fall of palsa mounds

from Peter D. Moore

THE peat bogs of high latitudes, where the mean annual air temperature is -1°C and the annual precipitation is less than 400 mm, are characterized by irregular peat-covered mounds with ice cores, called palsas¹. The interaction of the climatic, biological and geomorphological factors involved in the initiation and growth of these mounds has provided a rich area for speculation, but one in which experimental data have been conspicuously absent. Now, however, experiments on a bog in the north of Finland by Seppala^{2,4} have provided a firm basis for the understanding of palsa mound ecology.

Palsa mounds are normally 2–4 m high, though some may reach as high as 7 m in the Fennoscandian mires⁵. They can be seen to pass through a cycle of growth followed by collapse when the ice core melts and leaves a water-filled crater surrounded by a peaty rampart. Salmi has suggested that palsas began their existence during periods of unusually cold climatic conditions, and has provided palynological data from Finnish palsas, accompanied by radiocarbon dates, which indicate two periods of palsa formation, around 7,000 and 5,000 b.p. (ref. 6). But many palsa mires contain young developing palsa hummocks and the idea that initiation has been restricted to certain times in the past has not received general acceptance.

From their work on Canadian palsas, Raiton and Sparling have produced an

attractive theory of palsa growth and collapse based on the botanical succession which takes place on the elevated peat mass⁷. They found that as the ice core raises the peat surface to about 60 cm, changes in hydrological conditions encourage the invasion of small hummocks covered with *Sphagnum fuscum* by *Cladonia* lichens. They suggest that lichens have a considerable influence on the albedo of the mire surface; their white colour reflects such a high proportion of the incident radiation that the likelihood of summer melting of the ice core decreases and the palsa mound is allowed to grow more rapidly, aided by the high insulating properties of the drying peat blanket over its surface. Ultimately, the peat cover begins to erode as a result of wind action, and the disruption of surface vegetation exposes the dark-coloured peat with its low albedo. Summer absorption of radiation soon results in the rapid decay of the ice core and the collapse of the palsa.

The theory is very plausible and is supported by data on the albedo of the different vegetation types. But it lacks conviction, mainly in its explanation of the early stage of palsa formation simply as the development of a *Sphagnum* moss hummock. It is here that Seppala's proposals may fill the gap. Seppala has produced microclimatic data on the ground temperatures in the Arctic regions of Finland which show that frost

penetration is far greater where the snow cover is least³. He also finds that the basal ice that forms when frost penetrates to a depth of over 50 cm on a bog survives the following summer and provides a nucleus for further ice formation in the subsequent winter⁴. Those parts of a bog surface which are kept free of snow as a result of wind action might, therefore, provide promising locations for the initiation of new palsa mounds.

As an experimental test of his proposal, Seppala selected a flat unfrozen area of a palsa mire in northern Finland and installed temperature probes to a depth of 2 m into the peat at two sites. During the first winter, one site was experimentally cleared of snow on thirteen occasions, whereas on the control site the snow depth reached 90 cm. At the experimental site, frost penetrated 80 cm into the peat and, although the surface layers thawed during the following summer to a depth of 40 cm, the underlying 30 cm remained frozen throughout. At the control site, the winter frost only penetrated between 40 and 48 cm, not enough to maintain an ice core in the peat through the summer months. During the second winter, snow clearance was carried out eight times at the experimental site and frost penetration reached 1 m; a 50 cm thick ice core survived the following summer. The experiment continued for a third year and the ice continued to thicken.

The three years provided sufficient time for a palsa mound to begin to develop. The surface of the experimental site was 10 cm above the mire surface after one year, 20 cm after two years and 30 cm after three years. This was sufficient to produce some changes in vegetation, with the *Sphagnum* mosses and the cyperaceous plants becoming moribund as a result of summer drought. The surface peat was cracking, but there was no sign of lichen or shrub invasion. Interestingly, however, Seppala noted that the incipient palsa was lightly coloured, which could indicate that Raiton and Sparling's changes in albedo were commencing.

This experimental approach to the study of palsas has demonstrated that differential snow lie on the mire surface can result in the initiation of a palsa mound, and that the development of such a mound acts as a positive feedback to the process because it further decreases snow lie on the developing hummock. Vegetation changes then account for the further processes which result in the cyclical system of palsa growth and decay. □

1. Ruuhijarvi, R. *Ann. bot. Soc. fenn. Vanamo* **31**, 1 (1960).
2. Seppala, M. *Proc. 4th Can. Permafrost Conf.*, 36 (1982).
3. Seppala, M. *Oulanka Rep.* **4**, 20 (1983).
4. Seppala, M. *Permafrost: Proc. 4th int. Conf.*, 1127 (Washington, DC, 1983).
5. Ruuhijarvi, R. in *Ecosystems of the World. Mires: Swamp, Bog, Fen and Moor* Vol. 4B (ed. Gore, A.J.P.) 47 (1983).
6. Salmi, M. *Proc. 3rd int. Peat Conf., Quebec*, 182 (1968).
7. Raiton & Sparling *Can. J. Bot.* **51**, 1037 (1973).

Peter D. Moore is Senior Lecturer in the Department of Plant Sciences, King's College, 68 Half Moon Lane, London SE24 9JF.

Molecular genetics

RNA polymerase I promoters and transcription factors

from John Sommerville

EUKARYOTIC cell nuclei contain three types of RNA polymerase, each of which transcribes a distinct set of genes. Whereas RNA polymerase II transcribes thousands of different protein-coding genes and RNA polymerase III is engaged in the synthesis of several short RNA species, including 5S RNA and tRNA, RNA polymerase I acts almost exclusively on a single kind of transcription unit, the ribosomal RNA (rRNA) gene. For this reason, and because rRNA synthesis is a general feature of all cells, it might be expected that RNA polymerase I activity would be relatively simple to define in terms of the DNA sequences (promoters) and regulatory molecules (transcription factors) that determine the sites and efficiency of the binding of polymerase and of its initiation of transcription. In fact, only recently has significant progress been made towards identifying these components.

Promoters for polymerases II and III have been identified as conserved short sequences located at fixed distances from the transcription initiation site, but on comparing rRNA genes from yeast, *Drosophila*, *Xenopus*¹ and mouse², common sequences that could act as promoters are not obvious. Within more closely related groups, however, homologies can be seen (Fig. 1). For instance, the initiation regions of *Saccharomyces rosei* and *S. carlsbergensis* rRNA genes show greater than 90 per cent homology from nucleotide position -9 (relative to the transcription start point which is +1) to +64 and from +76 to beyond +125 (ref. 3); *Xenopus laevis*, *X. clivii* and *X. borealis* genes are

identical from position -9 to +4 (refs 4-6) and mouse, rat and human genes show strong homology from -1 to +18 (ref. 7). These comparative data suggest an important role for sequences located immediately around the initiation site.

Functional assays, however, implicate sequences that span a wider region. By using rDNA templates with various sequence deletions, it has been shown that correct and efficient initiation *in vitro* depends upon sequences in the region -43 to -27 in a *Drosophila* transcription system⁸ and -39 to -12 in a mouse system⁹. Recently, Muramatsu and co-workers¹⁰ have determined by deletion analysis that the sequence from -40 to +52 of mouse rDNA is adequate for full transcriptional activity, and the sequence from -22 to -2 is essential for minimum activity. These stretches of DNA contain two sequences that are conserved in mammals in addition to the sequence from -1 to +18 already mentioned: the sequence ATCTTT running downstream from position -38 and the sequence TATTG running downstream from position -20 (Fig. 1).

Some of the limitations in interpreting the effects of removal of whole blocks of sequence have been resolved by the construction of more precisely mutated templates. By introducing single and multiple base-pair changes into the -35 to -14 promoter region of mouse rDNA, Skinner *et al.*¹¹ have recently shown that conversion of C to T at positions -36, -27, -22, -21 and -13 has no effect on transcription efficiency, whereas the single transition of G to A at position -16 reduces the efficiency

of transcription by 95 per cent. This suggests a key role for the G at -16 in the interaction of a transcription factor with the mammalian rDNA promoter. Thus, the G at position -16, the TTT sequence around position -34 and the first few transcribed nucleotides^{10,12} may act together *in vivo*, like the TATA promoter for RNA polymerase II, to set the exact initiation point of transcription. As pointed out by Skinner *et al.*¹¹, the G at position -16 is conserved in higher eukaryotes, as is a T at position -1, so it will be interesting to assess the importance of these single nucleotides for rDNA transcription in other organisms.

But could there also be sequences that enhance transcription and lie further from the start site? Competitive effects between wild-type templates and deletion mutants indicate that sequences located between positions -169 and -45 influence the efficiency of *in vitro* transcription of mouse rDNA⁹. Also, in *Xenopus* rDNA there exist duplicated intact promoters, and 42 base pair repeats of the -114 to -72 region of the promoter, that extend several thousand base pairs upstream of the promoter closest to the gene^{13,14}. According to Reeder and co-workers¹⁵, these 42 base pair repeats (in either orientation) may bind transcription factors, which would explain why, in microinjected oocyte nuclei, rRNA genes with more repeats are preferentially transcribed. At limiting factor concentrations, rRNA genes would be either fully active or completely inactive, which indeed is what is observed during oocyte development. Yet, whatever their precise functions, at least some of the promoter elements so far identified appear to be conserved in sequence and/or location only within closely related groups.

The other side of the coin relates to the specificity of transcription factors. Successful *in vitro* transcription of an rDNA template by RNA polymerase I is dependent upon components that can be supplied from cell extracts. Grummt *et al.*¹⁶ showed that rDNA isolated from mouse, human and protozoan sources is transcribed accurately only if all the components are from the same species and concluded that transcription factors are species-specific.

The main approach to isolating and identifying transcription factors has been the chromatographic separation of cell homogenates. Mishima *et al.*¹⁷ have fractionated mammalian cell extracts into fractions A, B, C and D. Fraction C contains most of the RNA polymerase I and can be replaced by the purified enzyme; fraction B has a capacity to suppress random initiation and can be replaced by pure poly(ADP-ribose) polymerase; fraction A generally enhances transcriptional activity; but fraction D, which is indispensable for faithful initiation *in vitro*, has some degree of species-dependency in as much as mouse fraction D with human fractions A, B and C will not support transcription of human rDNA, but will support transcription of rat

	-40	-30	-20	-10	+1	+10
Mouse		ATCTTT		TATTG		TACTGACACGC
Rat		ATCTTT		TATTG		TACTGACACGC
Human		ATCTTT		TTTGG		TGCTGACACGC
<i>X. laevis</i>		CTACGCTTTT		GTGCGGGCAGGAAGGTAGGG		
<i>X. clivii</i>		CTACGCTTTT		GTGCCGACAGGAAGGTAGGG		
<i>X. borealis</i>		CTACGCTTTT		GTGCGGACAGGAAGGTAGGG		
<i>D. melanogaster</i>		TTT	G		TAGGT	
<i>T. pyriformis</i>			G		T	
<i>D. discoideum</i>			T	ATACATATA		
<i>P. polycephalum</i>			A	ATACATATA		
<i>S. carlsbergensis</i>			A	AGGTACTTCATGCGAAAGC		
<i>S. rosei</i>			G	AGGAACCTTCATGCGAAAGC		

Fig. 1 Major sequence homologies around the transcription initiation site of rRNA genes of mammals⁷, *X. laevis*⁴, *X. clivii*⁵, *X. borealis*⁶, *Drosophila melanogaster*⁸, *Tetrahymena pyriformis*¹⁸, *Dictyostelium discoideum*¹⁹, *Physarum polycephalum*²⁰, and *S. carlsbergensis* and *S. rosei*³. The conserved G at position -16 and the conserved T at position -1 are indicated (arrowheads) as is the start site and direction of transcription (arrow).

rDNA. Muramatsu is busy trying to identify the active component(s) in fraction D but reckons that he may need 1 kg of HeLa cells for this purpose. (By contrast, a single *Xenopus* oocyte yields 10 ng of a factor needed for transcription of 5S RNA genes by RNA polymerase III.)

Using a similar approach, S. Jacob of Hershey Medical Center (personal communication) has been able to separate to a high degree of purity transcription factors from rat mammary adenocarcinoma cells. When large quantities of cells, rather than isolated nuclei, are used as starting material, a complete transcriptionally active complex can be obtained. By dissecting out the components of this enzyme-factor complex a characterization of at least some factors should be possible.

Why should there be a lack of sequence conservation in RNA polymerase I promoters and species-dependency of transcription factors? As argued by Grummt *et al.*¹⁶, since RNA polymerase I transcribes only a single type of transcription unit, fixation of mutation within promoters of multiple rRNA genes can take place at a much higher rate than in polymerase II and III promoters, which would require the co-evolution of many different kinds of gene. The divergence of

rDNA promoters seems to have been, and perhaps had to be, accompanied by appropriate modification of the enzyme and/or transcription factors. What led to the evolutionary selection of the divergence will continue to puzzle us for some time. □

1. Long, E.O., Rebbert, M.L. & Dawid, I.B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1513 (1981).
2. Mishima, Y. *et al. Nucleic Acids Res.* **9**, 6773 (1981).
3. Verbeet, M. Ph. *et al. Gene* **23**, 53 (1983).
4. Boseley, P.G. *et al. Cell* **17**, 19 (1979).
5. Bach, R., Allet, B. & Crippa, M. *Nucleic Acids Res.* **9**, 5311 (1981).
6. McStay, B. & Bird, A. *Nucleic Acids Res.* **11**, 8167 (1983).
7. Financsek, I. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3092 (1982).
8. Kohorn, B.D. & Rae, P.M.M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3265 (1983).
9. Grummt, I. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6908 (1982).
10. Yamamoto, O. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 299 (1984).
11. Skinner, J.A., Ohrlin, A. & Grummt, I. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2137 (1984).
12. Learned, R.M. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 3558 (1983).
13. Moss, T. *Cell* **30**, 835 (1982).
14. Busby, S.J. & Reeder, R.H. *Cell* **34**, 989 (1983).
15. Reeder, R.H., Roan, J.G. & Dunaway, M. *Cell* **35**, 449 (1983).
16. Grummt, I., Roth, E. & Paule, M.R. *Nature* **296**, 173 (1982).
17. Mishima, Y. *et al. Nucleic Acids Res.* **10**, 6659 (1982).
18. Saiga, H. *et al. Nucleic Acids Res.* **10**, 4223 (1982).
19. Hoshikawa, Y., Iida, Y. & Iwabuchi, M. *Nucleic Acids Res.* **11**, 1725 (1983).
20. Blum, B. *et al. Nucleic Acids Res.* **11**, 8519 (1983).

John Sommerville is in the Department of Zoology, Bute Buildings, The University, St Andrews, Fife KY16 9TS.

Ecology

Oceanic noise and fish stocks

from Robert M. May

In recent years, many heavily exploited fish stocks have shown large fluctuations both in recruitment patterns and in overall abundance¹. Such fluctuations cannot, however, be unambiguously attributed to the effects of harvesting, because there is evidence that many pelagic fish species have undergone large and rapid changes in abundance in the past. Major changes in the abundance of Pacific sardine or Atlantic herring, for example, appear to have occurred every 50–100 years for the past several centuries; these fluctuations have taken place abruptly, and do not appear to be accompanied by similarly rapid changes in the physical environment. Recently, Steele and Henderson² have offered an explanation of this phenomenon that involves the interplay between a plausible model for the intrinsic dynamics of pelagic fish populations and the kind of environmental noise found in the oceans.

Most work in theoretical ecology that explores the effects of environmental noise upon population dynamics assumes the noise to be 'white'^{3,4}. For white noise, the variability manifested when one point in time is compared with another is independent of the time interval separating the two points; the variance is constant as the time interval increases from very short to very long times. This corresponds to the variance (per unit frequency) of the

Fourier-transformed frequency spectrum of the environmental noise being the same for all frequencies. In analogy with visible light, such a colourless noise spectrum is thus called white. There is, moreover, evidence that white noise may be an appropriate assumption for some kinds of variability in terrestrial environments: reviews of data for atmospheric variability and for tree rings show spectra that are white (at least down to frequencies corresponding to 100 year intervals)⁵.

Steele and Henderson argue that environmental fluctuations in the sea typically are different from those on land, exhibiting 'red' noise spectra in which the variance systematically increases towards the low-frequency end. In particular, for environmental factors such as temperature or sea level, for which relatively long records are available, the variance scales roughly as (frequency)⁻ⁿ, with *n* between 1.5 and 2; that is, the variance increases steadily as the time interval increases from days to decades⁶⁻⁸. Although Steele and Henderson rely mainly on empirical evidence, it is worth noting that simple stochastic models of the physical interactions between atmosphere and ocean also indicate that white noise in the atmosphere changes to red noise in the ocean⁹.

Steele and Henderson describe the population dynamics of their fish stocks by

a model with the following features: in the absence of predators, the population is regulated at relatively high density by the availability of resources; at high predation levels, the population is regulated at relatively low densities by predation; and there can be an intermediate regime in which the population has two alternative stable states (one essentially resource-limited, the other essentially predator-limited). Such models, with the possibility for two alternative equilibrium states, have already been discussed in relation to insect pests¹⁰, grazing herbivores¹¹, pathogens and other situations¹². The new twist imparted by Steele and Henderson is the insertion of environmental stochasticity with a red-noise spectrum into the system (specifically, into the predation term). The parameters in the model are estimated as those appropriate to pelagic fish, and the noise spectrum scales as (frequency)⁻², as above. Numerical studies with the model show the fish population undergoing moderate fluctuations about a steady high level, or about a steady low level, with abrupt transitions from one level to the other taking place every 50 years or so.

These observations could have important implications. In Steele and Henderson's words², "In the context of fisheries management, there is usually a tacit assumption that we are operating within periods where stocks are persistent, albeit with large year-to-year fluctuations in recruitment. The data ... suggest that the period 1920 to 1960, when much of fishery theory and practice was developed, was such an interval. More recently, large concurrent changes in stocks cast doubt on this assumption and raise critical management questions about the relative significance of environmental and fishery factors". Of course, the model explored by Steele and Henderson is a very simple one, and a good deal remains to be done before we accept that red noise in the ocean environment has tipped marine ecosystems from one state into another, largely independently of fishing. Nevertheless, as this work goes forward, it could well have relevance beyond fisheries. For example, insofar as marine environments differ from terrestrial ones in having predominantly red rather than white noise spectra, what would this imply for community structure and function? □

1. Lawton, J.H. & May R.M. *Nature* **309**, 744 (1984).
2. Steele J.H. & Henderson E.W. *Science* **224**, 985 (1984).
3. May, R.M. *Stability and Complexity in Model Ecosystems* (Princeton University Press, 1974).
4. Turelli, M. *Theor. Pop. Biol.* **12**, 140 (1977).
5. *Understanding Climatic Change: A Program for Action* (National Academy of Sciences, Washington DC, 1975).
6. Munk, W.H. & Cartwright, D.E. *Phil. Trans. R. Soc.* **259**, 1 (1966).
7. Wunsch, C. *Evolution of Physical Oceanography* (MIT Press, Cambridge, 1981).
8. Hays, J.D. *et al. Science* **194**, 1121 (1976).
9. Lemke, P. *Tellus* **29**, 385 (1977).
10. Ludwig, D., Jones, D.D. & Holling, C.S. *J. Anim. Ecol.* **47**, 315 (1978).
11. Noy-Meir, I. *J. Ecol.* **63**, 459 (1975).
12. May, R.M. *Nature* **269**, 471 (1977).

Robert M. May is Class of 1877 Professor of Zoology at Princeton University, New Jersey 08544.

Antineutrino astronomy and geophysics

Lawrence M. Krauss*, Sheldon L. Glashow† & David N. Schramm‡

* Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138, USA

† Department of Physics, Boston University, Boston, Massachusetts 02215, USA

‡ Department of Physics and Astrophysics, Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA

Radioactive decays inside the Earth produce antineutrinos that may be detectable at the surface. Their flux and spectrum contain important geophysical information. New detectors need to be developed, discriminating between sources of antineutrinos, including the cosmic-background. The latter can be related to the frequency of supernovae.

If there are more things in heaven and Earth than are dreamt of in our natural philosophy, it is partly because electromagnetic detection alone is inadequate. For sources which are visually obscured or which emit most of their energy in a form other than photons, new methods of detection must be developed. This has spurred the growth of neutrino astronomy, typified by the detection of neutrinos emitted in the interior of the Sun^{1,2} but although the usefulness of the weak interactions in probing astrophysical sources has been recognized, the potential of antineutrino detection has not been widely explored. That is our objective.

We demonstrate that the Earth is a rich source of antineutrinos whose detection can provide otherwise inaccessible key information on the internal structure and dynamics of the Earth. Moreover, by a consideration of the antineutrino background from other sources, we find that there may be a diffuse background of ~10-MeV antineutrinos from supernovae which, if detected, could yield information on the energy and frequency of supernovae.

That the Earth is a source of antineutrinos was first pointed out by Eders³ and Marx and co-workers⁴⁻⁶. Marx in particular was the first seriously to investigate the detection of terrestrial neutrinos and to recognize the importance of both resonant capture and inverse β decay of terrestrial antineutrinos in detection⁴. But a detailed analysis of the geophysical implications of the problems of measuring the actual flux of terrestrial antineutrinos has not been performed.

Earth as a source

Antineutrinos ($\bar{\nu}$) are produced in radioactive β decays of abundant isotopes inside the Earth as in

$$[A, Z] \rightarrow [A, Z+1] + e^- + \bar{\nu} \quad (1)$$

where A, Z are the atomic number and charge of the nucleus and e^- represents the β particle (electron). Isotopes which first decay by α emission yield neutron-rich products which subsequently undergo β decay. The discovery of these processes, fundamental in the study of geophysics, drastically altered estimates of the lifetime of the Earth based on the assumption that there is no substantial production of heat within the Earth⁷. Indeed, the energy released in these decays also to a large extent determines the dynamics of its interior.

The abundance of radioactive isotopes inside the Earth is thus of prime geophysical importance, but direct sampling is possible only at or near the surface⁸. Measurements of total heat flow out of the Earth, moreover, are uncertain because of the size of local variations and the inaccessibility of much of the Earth's surface⁸. But even if present estimates of the global heat

flux, centring on the value 40 TW^{8-10} , were much more accurate, the question of how much of the heat loss arises from production and how much from the fact that the Earth is still hot would remain unresolved. The difficulty is that heating from radioactive sources known in the surface layers is of the same order of magnitude as the total heat flux, so that even small abundances elsewhere would imply that the interior of the Earth is heating, not cooling.

Table 1 gives data for the lithosphere, which we define as the surface layer down to ~100 km depth containing the continental and sub-oceanic plates whose mass of $\sim 2 \times 10^{25}$ comprises about 1/300 of the Earth's total mass⁹. The mean abundances are given in Table 1 for the chief radioactive isotopes ^{40}K , ^{238}U , ^{232}Th , ^{87}Rb (ref. 11); actual abundances depend crucially on rock type and region⁸, generally decreasing from granite to basalts to dunites, with ^{238}U and ^{232}Th abundances varying by factors of 5–10. The abundance of ^{40}K is particularly sensitive to rock type and drops to almost zero in dunites⁸.

Our first objective is to calculate the spectrum of the antineutrino flux from the Earth, using the lithospheric abundances of radioactive materials as a starting point but within the constraints of what is known of the total heat flux. In reality, heat flux and the distribution of radioactive materials are not uniform over the surface of the Earth. Both are probably enhanced in continental crust¹² while heat flow from the oceans is probably largely due to plate formation there¹³. To begin with, though, we ignore this asymmetry.

Table 1 shows that the lithospheric ^{238}U and ^{232}Th decay chains account for an average heat production of ~17 TW, far exceeding that from ^{40}K and ^{87}Rb . Total radiogenic heat production at ~19 TW is approximately half the total heat flux at the surface.

In what follows, we calculate antineutrino fluxes from lithospheric radioactive isotopes only. For while most geophysical models¹³ suggest that concentrations of radioactive isotopes decrease rapidly at depths ≥ 20 km, even small concentrations in a mass 300 times as great as that of the lithosphere could powerfully affect antineutrino production. And while the total quantities of U and Th could not exceed twice those in the lithosphere without violating the constraints imposed by the total heat flux, K and Rb concentrations are not as constrained and, indeed, on some geophysical models, total K may be 10 times that in the lithosphere^{8,12,14}. Since essentially all antineutrinos produced within the Earth reach the surface without interacting, it follows that our estimates of surface flux must be underestimates, perhaps substantial.

From Table 1, the total antineutrino flux normal to the Earth's surface due to all decays in the lithosphere is of the order of $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ in an energy range up to 3.26 MeV, four orders of magnitude less than the solar neutrino flux of $10^{11} \text{ cm}^{-2} \text{ s}^{-1}$ for low-energy (<1 MeV) neutrinos. Above ~1 MeV, the spectral

† Permanent address: Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138, USA.

Table 1 Chief radioactive isotopes in the lithosphere ($m = 2 \times 10^{25}$ g)

Isotope	$t_{1/2} \times 1.44$ (s)	% Isotope abundance	Element abundance by weight	Flux $\bar{\nu}$ ($\text{cm}^{-2} \text{ s}^{-1}$)	Integrated flux	E_{max} (MeV)	Heat generation ($\text{cal g}^{-1} \text{ s}^{-1}$)	Total heat (W)
^{40}K	5.7×10^{16}	0.01	2.6×10^{-2}	3×10^6	1.1×10^7	1.31	7.2×10^{-13}	1.8×10^{12}
^{87}Rb	2.1×10^{18}	28	3.1×10^{-4}	1.3×10^6	4.6×10^6	0.274	1.8×10^{-13}	0.2×10^{12}
$^{232}\text{Th}^*$	6.3×10^{17}	100	1.9×10^{-5}	2.4×10^5	3.5×10^6	2.25	6.3×10^{-9}	7.3×10^{12}
				$\times 4$ $= 9.6 \times 10^5$				
$^{238}\text{U}^*$	2×10^{17}	99	4×10^{-6}	2.4×10^5	3.5×10^6	3.26	2.4×10^{-8}	9.6×10^{12}
				$\times 6$ $= 9.6 \times 10^5$				
			Total	6×10^6	2×10^7			19×10^{12}

* Data include decay chain.

region in which the terrestrial antineutrino intensity is greatest (see below), the solar flux is of the order of $\sim 10^8 \text{ cm}^{-2} \text{ s}^{-1}$. Thus in this region of the spectrum, we may expect the terrestrial signal to be perhaps one order of magnitude and not four weaker than that due to solar neutrinos. Geometry further enhances the terrestrial signal, as will be seen, if the source distribution is a shell near the surface.

The quantity relevant for antineutrino measurement, which we call $F_{\bar{\nu}}$, is the antineutrino flux at the surface integrated over all directions. This can be expressed as a radial integral in terms of the antineutrino production rate per unit volume $n_{\bar{\nu}}(r)$, which is directly related to the distribution of radioactive sources. If R is the radius of the Earth,

$$F_{\bar{\nu}} = \frac{1}{2R} \int_0^R n_{\bar{\nu}}(r) r \ln \frac{1+r/R}{1-r/R} dr \quad (2)$$

If $n_{\bar{\nu}}(r)$ is constant in a shell going from the surface to some smaller radius s , equation (2) can be integrated exactly to yield:

$$F_{\bar{\nu}} = \frac{3}{2} \frac{N_{\bar{\nu}}}{4\pi R^2(1-s^3/R^3)} \times \left[(1-s/R) \left[1 + \frac{1}{2} \left[1 + s/R \right] \log \left[\frac{1+s/R}{1-s/R} \right] \right] \right] \quad (3)$$

where $N_{\bar{\nu}}$ is the total rate of antineutrino emission from the Earth. This function varies from $3.5 (N_{\bar{\nu}}/4\pi R^2)$ for a shell of depth ~ 30 km, to $1.5 (N_{\bar{\nu}}/4\pi R^2)$ for a uniform distribution throughout the Earth, to $1.0 (N_{\bar{\nu}}/4\pi R^2)$ for a distribution concentrated at the Earth's centre. For this last distribution, all the flux is normal to the Earth's surface and thus the integrated flux is equal to the normal flux. For all other distributions, integrated fluxes are enhanced relative to normal flux estimates. However, surface distributions (like the ones we consider here) are only enhanced by a factor ~ 2 compared with uniform distributions, implying that sensitivity to significant abundances deep inside the Earth is not lost. (We use a conservative estimate of 30 km as the shell size here. A smaller value could increase the base signal from the lithosphere as can be seen from equation (3).)

This is taken into account in both Table 1 and Fig. 1, which presents the total flux ($F_{\bar{\nu}}$) per unit energy of antineutrinos from the lithosphere, calculated using an appropriately normalized allowed β -decay spectrum for ^{40}K , ^{87}Rb and each of the β^- emitters in the U and Th chains. Figure 1 shows that there are several well defined peaks and plateaus in the terrestrial antineutrino spectrum related to decays of the four chief isotopes of Table 1. Figure 2 shows the expected antineutrino/neutrino spectrum at the Earth from all sources over a wide range of energies including the solar and terrestrial peaks, black-body neutrinos and antineutrinos, and a background of antineutrino from supernovae.

Essentially all antineutrinos produced inside the Earth reach the surface without interaction. Thus their measurement would be an ideal probe of the Earth's interior and be useful for studying regions inaccessible by other means. With direct sampling we can probe at best the upper mantle region in contact

with the lithosphere using the outflow from volcanoes. Modulo small enhancement factors for surface distributions, however, the antineutrino signal gives a whole Earth measurement. Subtracting known contributions from the lithosphere gives a direct measurement of the internal properties of the Earth. Important geophysical information which can, in principle, be obtained by measuring the antineutrino flux from different radioactive isotopes is given below.

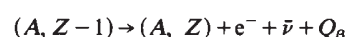
(1) **U, Th.** Uranium and thorium decays are responsible for most of the heat production inside the Earth. The associated antineutrino flux, therefore, gives a direct measurement of heat production inside the Earth due to radioactivity. These antineutrinos are produced only in radioactive decays and are not related to any other heat source, or to residual heating. Thus we can determine that component of heat flow produced from radioactivity. This both allows a better determination of whether the Earth is cooling and provides energetic data for understanding the Earth's dynamics. Also, determination of internal abundances of U and Th imparts information on the geology of the Earth's mantle and core—in particular to check the assumption that this radioactivity is concentrated on the surface, and also in the continental crust⁹.

(2) **^{40}K .** Table 1 shows that ^{40}K decays produce the largest overall flux from the lithosphere. The abundance of ^{40}K is sensitive to geology, and is assumed to be zero in dunites and in the mantle and core. However, as its abundance is not constrained by heat flow considerations, any amount of ^{40}K in the interior could increase the flux due to ^{40}K decays by perhaps a factor of 10 (ref. 14). The sensitivity of ^{40}K to rock type means that a measurement of such a large flux could provide important data on the composition of the mantle and core.

(3) **^{87}Rb .** Measurement of the low-energy antineutrinos from ^{87}Rb can have important implications for our understanding of convection currents in the upper mantle. These currents are assumed to be of the order of 700 km in depth⁹ and during solidification of the crust transported Rb from the upper mantle to the crust. How much the abundance of ^{87}Rb in the crust differs from its whole Earth value thus gives an indication of how much material has been deposited in the crust, and thus of how deep this upper mantle is⁹. This convection layer is also related to the dynamics of heat storage. Determining the amount of internal heat generated by radioactivity can also shed light on how much is being stored, and hence on the size and dynamics of the outer convection layer.

Detecting low-energy antineutrinos

Antineutrinos with energies in the MeV range interact with nuclei producing radiochemical transitions from (A, Z) to $(A, Z-1)$ atoms through two different weak interaction processes related to the inverse of the β -decay process (see equation (1)):



where Q_{β} represents the energy release beyond the rest energy

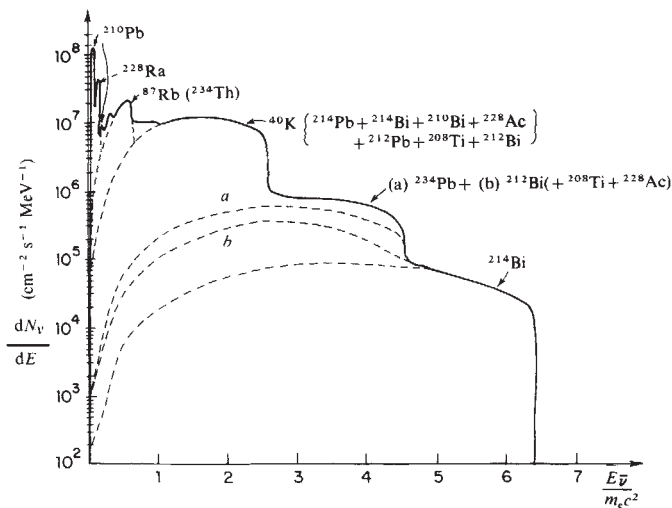


Fig. 1 Number density of antineutrinos at the Earth's surface due to abundance of chief radioactive isotopes in the lithosphere.

of the electron. By interchanging particles in the initial and final state we can consider the reactions:

$$(A, Z-1) \leftarrow (A, Z) + e^- + \bar{\nu} + Q_\beta \quad (4a)$$

$$e^+ + (A, Z-1) \leftarrow (A, Z) + \bar{\nu} + Q_\beta + 1.02 \text{ MeV} \quad (4b)$$

Note their differing kinematics. Process (4a) (resonant orbital electron capture¹⁵) has threshold energy identical to the related β -decay process, but occurs as a resonant process only for incoming antineutrinos in a small range around this energy. Process (4b) (inverse β decay), on the other hand, requires an additional 1.02 MeV of energy for the incoming antineutrino, but can occur for all energies above this threshold. Thus these processes can be exploited to probe different energy regions of the terrestrial antineutrino spectrum (Fig. 1).

In addition, each of the two processes predominates in different regions of the (A, Z) space of target atoms. The cross-section for the resonant process is given by¹⁵:

$$\sigma_{\text{res}} \sim S \frac{1.66 \times 10^{-40}}{(137)^3} |\psi(R_0)|^2 \frac{\rho(E_{\text{res}})}{ft} \text{ cm}^2 \quad (5)$$

where $|\psi(R_0)|^2$ is the probability density for the electron at the nuclear surface, ft is the 'reduced lifetime' characterizing the nuclear matrix element for the related β decay, $\rho(E_{\text{res}})$ is the density of antineutrinos per unit energy interval at the resonant energy, and S is a spin-statistics factor for the final state. As $|\psi(R_0)|^2$ increases with increasing Z , equation (5) is maximized for detectors with high Z , and low ft values (allowed transitions).

Note that the dependence of σ_{res} on $\rho(E_{\text{res}})$ implies an enhanced interaction rate for incoming antineutrino distributions which are sharply peaked in the region of the resonance. In particular, low-energy sources such as ^{87}Rb can have large $\rho(E)$ values (see Fig. 1) without extremely large integrated flux densities. To obtain Fig. 1, we used $\rho(E)$ for allowed β transitions, including the Coulomb correction factor which biases the antineutrino spectrum towards higher energies¹⁶, and using known ft and Q_β values for the relevant terrestrial isotopes. As terrestrial decays are not of the allowed type, the actual $\rho(E)$ would be slightly modified by factors enhancing the ends of the spectrum¹⁷. We approximated these effects by normalization of $\rho(E)$.

The inverse β -decay cross-section is energy dependent and is given by¹⁶:

$$\sigma_{\nu}(E) = S \frac{2.6 \times 10^{-41}}{ft} F(Z, \epsilon') (\sqrt{\epsilon'^2 - 1}) \epsilon' \quad (6)$$

where $\epsilon' = E - Q_\beta - 1$ is the energy of the emitted positron (in units of the electron mass), and $F(Z, \epsilon')$ is the Coulomb factor

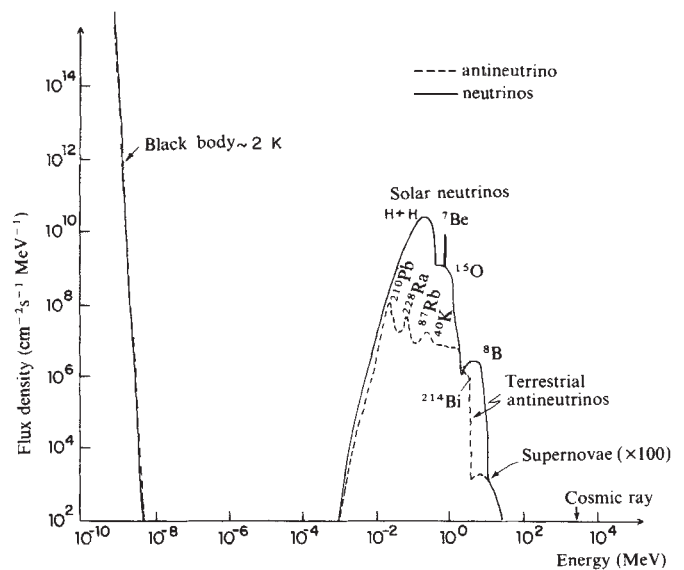


Fig. 2 Spectrum of neutrinos and antineutrinos at the Earth's surface (continuous sources in $\text{cm}^{-2} \text{s}^{-1} \text{MeV}^{-1}$, line sources in $\text{cm}^{-2} \text{s}^{-1}$. Supernovae peak is enhanced by a factor of ~ 100).

which tends in this case to suppress the cross-section near threshold with increasing Z . The total interaction rate is obtained by integrating equation (6) over the range of available energies, weighted by the factor $\rho(E)$.

Thus, process (4a) increases for higher detectors, while process (4b) decreases, and (4a) is favoured for narrow spectra, while (4b) is more sensitive to broader spectra going to higher energies.

These differences can be exploited in detector design to maximize sensitivity to different antineutrino sources. Low Z detectors will tend to involve inverse β -decay interactions which will be sensitive to the higher-energy terrestrial β decays of ^{40}K , ^{238}U , ^{232}Th , while high Z detectors can, through resonant capture, pick out peaks in the spectrum, such as those at low energy for ^{87}Rb and ^{212}Pb decay. We now consider two model detectors, one at high Z and one at low Z ; these were chosen without regard to actual experiments. Table 2 shows the calculated interaction rates for the detectors ^3He and ^{209}Bi . These results justify our expectations.

Finally, before considering the practicality of antineutrino detection, we consider whether the total interaction rates of Table 2 may be increased significantly using different detectors. First, for inverse β decay we recall that detection efficiency is maximized for detectors with low Z , low ft values and low Q_β . ^3He is the lowest Z stable product of a β -decay except for a free proton, and Q_β for the transition, $^3\text{H} \rightarrow ^3\text{He}$ is even lower than the $n \rightarrow p$ transition. Also, as $\log ft \approx 3$ with equality for transitions which are both allowed and favoured, and thus have maximal nuclear wave function overlap, ^3He is the optimal inverse β -decay type detector. Thus the maximal possible interaction rate due to antineutrinos from the lithosphere along which interact through inverse β decay is of the order of 0.4 TAU (SNU). Of course, this number may increase by up to an order of magnitude due to antineutrinos from the rest of the Earth.

The maximum resonant capture rate is more difficult to determine. The rate of 0.2 TAU in the ^{209}Bi model detector is about the same order as the maximum inverse β decay in the He detector. ^{209}Bi is the highest Z stable product of a β^- decay, and hence optimizes this aspect of resonant capture. However, the ft value of this detector ($\log ft = 5.5$) is over two orders of magnitude greater than that of the He detector. Since $\sigma \sim (ft)^{-1}$, a high Z target with $\log ft \sim 3$ would have a detection rate of $\sim 10^2$ TAU which is not possible.

The table of isotopes indicates that there is no ground state-to-ground state transition with $ft \sim 10^3$. We might, however, expect

Table 2 Two model detectors

		${}^3_2\text{He} \rightarrow {}^3_1\text{H}$	H_1 lifetime = 12.3 yr $\log ft = 3.1$ $Q_{\text{B}} = 0.01861 \text{ MeV}$		${}^{209}\text{Bi} \rightarrow {}^{209}\text{Pb}$	Pb lifetime = 3.3 h $\log ft = 5.5$ $Q_{\text{B}} = 0.64$	
		Rates		Rates			
Terrestrial process	E_{ν} max	Resonant (TAU)	Inverse (TAU)	Resonant (TAU)	Inverse (TAU)		
${}^{40}\text{K}$	1.31	1.8×10^{-9}	0.11	8.5×10^{-2}	0		
${}^{87}\text{Rb}$	0.274	3.1×10^{-7}	0	0	0		
${}^{238}\text{U}$							
${}^{234}\text{Th}$	0.191	5.2×10^{-9}	0	0	0		
${}^{234}\text{Pa}$	2.29	2.3×10^{-9}	0.11	10^{-4}	1.2×10^{-5}		
${}^{214}\text{Pb}$	1.03	1.4×10^{-9}	0	8.5×10^{-2}	0		
${}^{214}\text{Bi}$	3.26	4×10^{-11}	5.3×10^{-2}	2.8×10^{-3}	9×10^{-6}		
${}^{218}\text{Pb}$	0.061	9×10^{-8}	0	0	0		
${}^{210}\text{Bi}$	1.16	1.2×10^{-9}	1.4×10^{-2}	1.1×10^{-2}	0		
${}^{232}\text{Th}$							
${}^{228}\text{Ra}$	0.055	2.5×10^{-5}	0	0	0		
${}^{278}\text{Ac}$	2.18		1.4×10^{-2}	6×10^{-3}	10^{-7}		
${}^{212}\text{Pb}$	0.58	$0(10^{-9} - 10^{-11})$	0	0	0		
${}^{212}\text{Bi}$	2.25		3.1×10^{-2}	10^{-5}	6×10^{-6}		
${}^{208}\text{Tl}$	1.80		2.8×10^{-2}	10^{-4}	2×10^{-6}		
Total from lithosphere		2.5×10^{-5}	0.36	0.2	2.6×10^{-5}		

1 TAU = 10^{-36} interactions per target atom s^{-1} (and is equivalent to 1 SNU).

that a transition from an (A, Z) nuclei to an excited state of some $(A, Z-1)$ nuclei may have much better wave function overlap producing a minimal $\log ft$ value. This is not possible because such a value occurs when the proton making the transition does not change nuclear levels. For all stable high Z nuclei, $n_{\text{neutron}} \gg n_{\text{proton}}$, and therefore all proton-to-neutron transitions must result in level changes, implying at best a $\log ft \geq 4.5$ (ref. 16).

Thus transitions to excited states cannot significantly increase resonant cross-sections compared with that of ${}^{209}\text{Bi}$, and if $Z < 83$ for such detectors then any increase due to a slightly reduced ft value is largely compensated by the Z dependence of the cross-section. Some gains can be made if $Q_B < 0.274 \text{ MeV}$ so that the detector is sensitive to low-energy antineutrinos since $\rho(E)$ for these antineutrinos is greater by at least a factor of 2 (see Fig. 1) compared with those in the range of the ${}^{209}\text{Bi}$ detector. In general, however, for both the inverse β decay and resonant capture processes, the total interaction rate of antineutrinos due to decays in the upper lithosphere is always $\leq 0.5 \text{ TAU}$.

(Antineutrino interactions may also be measurable by non-radiochemical means, such as water type targets using photo-detection of γ rays associated with antineutrino scattering¹⁷. However, these cannot distinguish the antineutrino signal from comparable or larger signals because of other background effects, and solar neutrinos.)

An interaction rate of 0.5 TAU translates into an event rate of $\sim (1 \text{ AMU}/M_{\text{nucleus}}) \text{ events h}^{-1} \text{ kton}^{-1}$ of detection material. Whether this is acceptable depends on the feasibility of amassing 1 kton of reasonable detection material, and then extracting the signal from the background. For comparison, we consider closely related solar neutrino experiments.

The Brookhaven solar neutrino experiment utilizes the reaction $\nu + {}^{37}\text{Cl} \rightarrow {}^{37}\text{Ar} + e^-$, which is the inverse of the electron capture decay of ${}^{37}\text{Ar}$, with a half life of 35 days (ref. 1) and a threshold of 0.81 MeV. ${}^{37}\text{Cl}$, with an isotopic abundance of 25% is easily obtainable. A 400,000-l (0.6 kton) tank of perchlorethylene, C_2Cl_4 , was used as a detector and placed underground in the Homestake Gold mine at a depth of $\sim 1,500 \text{ m}$ to reduce cosmic-ray production of ${}^{37}\text{Ar}$. The ${}^{37}\text{Ar}$ produced (along with a little ${}^{36}\text{Ar}$ placed in the tank to check recovery efficiency), being a noble gas is easily removed from the liquid by purging with helium gas. After its separation from the helium, its radioactive decay is measured to determine the net production of ${}^{37}\text{Ar}$. A production rate of $\sim 2.0 \pm 0.3 \text{ SNU}$ ($= \text{TAU}$) is obtained, with

a background from known sources of $\sim 0.35 \text{ SNU}$ (ref. 1). The actual ${}^{37}\text{Ar}$ production rate was ~ 1 atom per 2 days.

To detect antineutrino event rates of the order of ~ 1 per day the following aspects of the ${}^{37}\text{Cl}$ experiment must be reproduced.

(1) To obtain ~ 1 event per day due to 0.5 TAU (at best), of the order of 1 kton of material must be available. The material must be cheap to produce, or to borrow (as it is not destroyed in the detection process).

(2) Because of the extremely small number of atoms produced, separation techniques must be almost 100% efficient. This is possible for reaction products which are noble elements. If this is not the case, more sophisticated chemical separation must be possible. Also, use of a liquid or gas detector for separation purposes seems most feasible.

(3) If radiochemical detection is envisioned the reaction product must have a lifetime long enough for significant production to occur but short enough so that subsequent decays can be measured—of the order of $1 \text{ day} < \tau < 1 \text{ yr}$.

The model targets discussed demonstrate the maximum possible detection rates due to the two interaction processes. They were not selected for their practicality. Actual targets may have detection rates which are significantly less than these maximum values. Table 3a gives a complete list of all stable target materials with $A_Z \rightarrow A_{Z-1}$ transitions induced by antineutrinos, and with product lifetimes $> 1 \text{ day}$.

Table 3 shows that there are no allowed ground-state transitions which involve noble gas products. Indeed, the only such process, $\text{Rb} \rightarrow \text{Kr}$, is strongly suppressed both in ground-state and excited-state transitions. This implies that the proven separation technology from the ${}^{37}\text{Cl}$ solar neutrino experiment cannot be used in antineutrino detection. Other chemical separation techniques are possible, however, especially if the chemistry of target and product differ significantly as in the case of Cl or Br , whose products can undergo reduction reactions and subsequent water layer separation (R. Davis Jr, personal communication).

Next we observe that all ground-state transition rates are at least an order of magnitude smaller than the ${}^3\text{He} \rightarrow {}^3\text{H}$ reaction. In particular the ${}^{35}\text{Cl}$ reaction which might be considered because it could utilize the same tank and materials as the solar neutrino experiment and has reasonable chemistry for separation, would only result in a transition rate of $\sim 1 \text{ atom yr}^{-1}$. ${}^3\text{He}$ is unlikely to be available in large enough quantities and thus we must consider measuring rates of $10^{-1} - 10^{-2} \text{ TAU}$, or else utilizing transitions to excited states. As the $\text{Zn} \rightarrow {}^{64}\text{Ni}$ reaction which seems to be the next favoured ground-state transition

Table 3 Antineutrino targets with product lifetime ≥ 1 day and $Q_{ns} < 2$ MeV (in order of increasing $\log ft$)

Target process	Q_B (MeV) (ground state transition)	$\log ft$	Product lifetime	Interaction rate (TAU)	Sensitivity to terrestrial decay
$^3\text{He} \rightarrow ^3\text{H}$	0.0186	3.1	12 yr	0.32	^{238}U , ^{40}K , ^{232}Th
$^{33}\text{S} \rightarrow ^{33}\text{P}$	0.248	5.0	25 day	0.002	^{232}Th , ^{238}U , ^{87}Rb
$^{35}\text{Cl} \rightarrow ^{35}\text{S}$	0.167	5.0	88 day	0.004	^{232}Th , ^{238}U , ^{87}Rb
$^{121}\text{Sb} \rightarrow ^{121}\text{Sn}$	0.383	5.0	27 h	0.01	^{40}K , ^{238}U
$^{64}\text{Zn} \rightarrow ^{64}\text{Cu} \rightarrow ^{64}\text{Ni}$	0.573	5.3	Stable	0.02	^{40}K , ^{238}U , ^{232}Th
$^{77}\text{Se} \rightarrow ^{77}\text{As}$	0.68	5.7	1.6 day		
$^{175}\text{Lu} \rightarrow ^{175}\text{Vb}$	0.467	6.3	5.5 day	$\sim 10^{-2}$	
$^{169}\text{Tm} \rightarrow ^{169}\text{Er}$	0.34	6.4	9.4 day		
$^{185}\text{Re} \rightarrow ^{185}\text{W}$	0.49	6.5	75 day		
$^{177}\text{Hf} \rightarrow ^{177}\text{Lu}$	0.497	6.6	6.7 day		
$^{63}\text{Cu} \rightarrow ^{63}\text{Ni}$	0.067	6.7	92 yr	0.04* (0.15–0.2 MeV)	^{87}Rb , ^{232}Th
$^{153}\text{Eu} \rightarrow ^{153}\text{Sm}$	0.801	7.2	2 day		
$^{199}\text{Hg} \rightarrow ^{199}\text{Au}$	0.46	7.5	3.15 day		
$^{151}\text{Eu} \rightarrow ^{151}\text{Sm}$	0.076	7.6	87 yr	0.15 (0.15–0.23 MeV)	^{87}Rb , ^{232}Th
$^{141}\text{Pr} \rightarrow ^{141}\text{Ce}$	0.58	7.7	33 day		
$^{161}\text{Dy} \rightarrow ^{161}\text{Tb}$	0.58	7.8	7 day		
$^{155}\text{Gd} \rightarrow ^{155}\text{Eu}$	0.248	8.2	1.8 yr		
$^{14}\text{N} \rightarrow ^{14}\text{C}$	0.156	9.0	5×10^3 yr		
$^{85}\text{Rb} \rightarrow ^{85}\text{Kr}$	0.67	9.1	10.7 yr		
$^{113}\text{In} \rightarrow ^{113}\text{Cd}$	0.58	9.2	14 yr		
$^{39}\text{K} \rightarrow ^{39}\text{Ar}$	0.565	9.9	269 yr		
$^{204}\text{Pb} \rightarrow ^{204}\text{Tl}$	0.765	9.9	3.8 yr		
$^{79}\text{Br} \rightarrow ^{79}\text{Se}$	0.154	10.8	6.5×10^4 yr	0.08 (0.25 MeV)	^{87}Rb , ^{232}Th
$^{107}\text{Ag} \rightarrow ^{107}\text{Pd}$	0.035	10.8	7×10^6 yr		

Estimate excited state transition interaction assuming $\log ft = 5.0$ for allowed transition. Excited state Q_B given in parentheses.

after He would result in an event rate of only ~ 1 atom per month per kton, the second possibility seems appropriate unless detectors larger by at least an order of magnitude than the ^{37}Cl detector are envisioned.

Rates for transitions from ground state to excited state, if they are allowed, can, for high Z atoms, approach or equal the ^{209}Bi rate of ~ 0.2 SNU. This can lead to event rates of the order of 1 per day. However, because the radioactive lifetime of many of product atoms (once they have relaxed to their ground state) is too long to use radiochemical detectors we must consider new technology for separation of the product atoms and for their subsequent detection. However, new methods, such as resonant ion spectroscopy, make detection of small number of 'stable' atoms possible. Current sensitivity for resonant ion counting is about 100–1,000 atoms. Such values can, in principle, be produced with running times of the order of a year. Note that proposals exist for proton decay detectors which will measure stable decay product atoms¹⁸. A proton lifetime of $\sim 10^{31}$ yr corresponds to a rate of $\sim 3 \times 10^{-3}$ TAU. For long-lived transition products, the possibility of archeological detection might be considered by searching for products which can have accumulated over long periods in ordinary rock. However, such abundances would be at best of the order of $\sim 10^{-19}$.

The experimental situation with regard to terrestrial antineutrino detection today is similar to that at the time when the first experiments for solar neutrinos were designed. Large-scale machines using new technology will be required, probably exploiting excited state transitions (possibly ^{79}Br) with rates at best ~ 0.1 – 0.2 TAU and new resonant ion spectroscopy techniques.

Other aspects of the experimental situation are encouraging. First, such detectors will be able to be calibrated using nuclear reactor antineutrinos: this was not possible for solar neutrino detectors. Such calibration is important both for terrestrial antineutrino measurement and for checking cross-sections of low-energy antineutrinos on complex nuclei which have never been directly measured.

Second, terrestrial antineutrino fluxes have a known baseline—that due to known levels of radioactivity in the lithosphere—whereas ^{37}Cl solar neutrino fluxes depend on processes

and abundances in the Sun which are not well known, and for which a lower limit is not easy to establish. If neutrino oscillations¹⁹ occur (with three neutrino species) over length scales that are small compared with the radioactive shell depth, the lithospheric signal could be reduced by a factor of 3 below the expected baseline level. While many effects can increase the detection rate, if sensitivities allow an observation of such a possible reduction, this would point to neutrino oscillations.

Even a null result can be significant if the detector sensitivity were of the order of the expected lithosphere flux. As stated earlier, abundances of ^{40}K and ^{87}Rb can be significantly above those present in the lithosphere—by up to a factor of 10. Such abundances can be ruled out by null results, allowing empirical tests of some geophysical models. Finally, as excited state detectors have long-lived final products, longer running periods and hence increased sensitivity are possible.

Whether such first generation terrestrial antineutrino detectors would yield more information than similar solar neutrino detectors depends on the possible backgrounds. These we separate into four types: nonantineutrino-induced events, antineutrinos from other terrestrial sources and from local radioactive deposits, and extraterrestrial antineutrino backgrounds. The extraterrestrial background is discussed below.

Unless the background for the terrestrial antineutrino experiment is less than that of the known background in the ^{37}Cl solar neutrino experiment (~ 0.3 SNU) it will overwhelm the signal. The background for both experiments originates from cosmic ray muons. In the case of solar neutrino detectors, secondary protons can mimic ν induced reactions through a (p, n) type replacement. For antineutrino detection, however, it is either tertiary neutrons (from spallation reactions), or stopped muons themselves which can mimic the terrestrial antineutrino signal—in the first case by (n, p) type interactions, or in the second through an inverse β -decay type process involving muon capture.

We believe the second background is more problematic than the first. Deep underground, average surviving muon energies will be of the order of ~ 300 GeV (ref. 20). Most spallation neutrons created in the detector will not lose sufficient energy in the detector volume to result in $A_Z \rightarrow A_{Z-1}(n, p)$ reactions

which do not further disrupt target atoms. Also, slow neutrons from the surrounding rock outside the detector (from local radioactivity as well as spallation effects) can be absorbed if suitable shields are placed around the detector. Thus we concentrate our quantitative analysis on the case of stopped muons.

At the depth of the Brookhaven CI detector¹⁷, muons are stopped at an average rate of $2 \times 10^{-3} \text{ ton}^{-1} \text{ day}^{-1}$. Whether these are captured and mimic antineutrino-induced transitions depends both on the capture probability and the probability that no neutrons are emitted from the excited nucleus after muon capture²¹. The first probability increases strongly with Z , while the second is generally small ($<10\%$) (see ref. 21). Also, using the fact that the stopping rate is a function of density, we find that for high Z targets the muon background at these depths should be about equal to the antineutrino signal for excited-state detectors with detection rates of $\sim 0.2 \text{ TAU}$. For low Z targets it is $<0.2 \text{ TAU}$. However, the only efficient detector at low Z is the ^3He . Thus we consider three possible ways to reduce this background. First, at only slightly greater depths the stopped muon rate drops asymptotically to $\sim 1/10$ their value at the Davis experiment depth¹⁷. Then, one might dope the detector with high Z efficient muon capturing material to reduce to number captured by desirable target atoms. A combination of these techniques may reduce this background to 10–20% of the antineutrino signal. One could determine quantitatively both this background, and that due to neutrons by using radiochemical detectors with Q_β values $\geq 2 \text{ MeV}$ which are insensitive to terrestrial antineutrinos but will react with neutrons and stopped muons.

Antineutrinos from other terrestrial sources consist primarily of atmospheric antineutrinos from cosmic-ray muon and pions, but we have also considered possible backgrounds from sources such as proton decay and nuclear weapons testing. Extrapolating fluxes of atmospheric neutrinos and antineutrinos estimated by other authors^{22–24}, to low energies, we estimate a maximum flux density of the order of $10^{-3} \text{ cm}^{-2} \text{ s}^{-1} \text{ MeV}^{-1}$ in the 1–10 MeV range. This background, even if it were to remain constant up to 50–100 MeV, would give $<1\%$ effect, both for resonant capture type detectors sensitive only to low-energy antineutrinos, and for inverse β -decay detectors which are also sensitive up to higher energies. With regard to possible antineutrino backgrounds from proton decay, with a proton lifetime of $>10^{31} \text{ yr}$ (ref. 25), even if the dominant decay mode is to antineutrinos the maximum flux is of the order of $10^{-4} \text{ cm}^{-2} \text{ s}^{-1}$ and is again negligible. Nuclear weapons explosions, on the other hand, can yield pulses of antineutrinos from the decay of fission products. However, a 1 Mton explosion at 1,000 miles from the detector will yield a maximum pulse of 10^{10} antineutrinos which, while momentarily swamping the terrestrial flux, yields a signal which is equivalent to that obtained in only 1 h of running time in the constant flux due to radioactivity. Thus unless the detector is built very close to a test site this background is also negligible. (Similar results apply for backgrounds from reactors.)

In discussing signal enhancement from local radioactivity, we assume that radioactive distributions are uniform in a shell $\sim 30 \text{ km}$ in depth from the surface, with an average concentration so that total abundances in the lithosphere agree with known values. Except for an enhancement factor (~ 3.5), the flux through the surface can be written

$$F_{\bar{\nu}} = \frac{C_{av} V_{\text{shell}}}{4\pi R_c^2} = \frac{C_{av} \frac{4}{3}\pi [R_c^3 - (R_c - d)^3]}{4\pi R_c^2} \quad (7)$$

where C_{av} is the average lithosphere source concentration, V the volume of the surface shell, $d = 30 \text{ km}$ (depth of shell). Now since $d \ll R$ we can expand equation (7) to obtain in first order:

$$F_{\bar{\nu}} = C_{av} d \quad (8)$$

Hence the relevant scale in the problem is the shell thickness d .

Thus, modulo geometric factors, we need only consider the quantity $F_{\bar{\nu}}' \approx C_{\text{local}} d_{\text{local}}$, where C_{local} is the enhanced con-

centration of a local source of scale size d_{local} , to see if local sources can overwhelm the global signal.

Whether these conditions are satisfied depends on the actual distribution of radioactive ore. For example, the minimal allowable C_{ore}/C_{av} for uranium to be minable is $\sim 10^3$ (refs 26, 27). But the largest single deposit in the United States, Jackpile, contains an ore body some 1,300 ft wide, 2,000 ft long, 20 ft in average thickness, ranging from surface outcroppings to a depth of 2,000 ft (ref. 27). Even assuming an optimal surface layer, we get $F_{\bar{\nu}\text{local}} = 5 \bar{\nu}_{\text{lithosphere}}$ for a detector directly above the distribution, which is too small a signal (of the order of 1 TAU) to make such prospecting practical.

Larger-scale enhancement of radioactive abundances may be of more geological interest, however. Indeed, the continental crust is expected to have an enhanced concentration of radioactive isotopes compared with the lithospheres as a whole¹³. Such enhancement could increase the measured signal compared with our estimates based on a spherically symmetric distribution. (Also, for constant total radioactivity, a smaller shell depth than we assumed could increase the signal compared with our initial estimates.) In general, however, as far as initial global measurements are concerned, while large-scale local region $>0(30 \text{ km})$ of enhanced deposits exist²⁸, only the major deposits, which can easily be avoided, might yield signals significantly greater than the mean global signal.

However, regions exist where it is expected that radioactive abundances may be significantly less than average. For example, heat flow under the oceans has been modelled as being due to convective plate formation, rather than radioactivity¹³. Some empirical confirmation of this large-scale anisotropy might be possible if detectors with sensitivity below the expected lithospheric level were placed in the ocean.

Thus, the only background which may pose difficulty for extracting the large-scale terrestrial antineutrino signal is that due to cosmic-ray muons, and we believe this can be reduced to an acceptable level in a way which can be checked independently.

Extraterrestrial antineutrinos

Except for rare nearby supernovae pulses, the solar neutrino flux masks all other possible extraterrestrial sources in the 0.1–15 MeV range. The terrestrial antineutrino flux is both smaller and more localized in energy, and will we believe, make antineutrino rather than neutrino detectors more useful for astrophysical purposes. In particular, a diffuse background of antineutrinos (and neutrinos) from supernovae exists which, if detectable, can yield information on supernova frequency in galaxies. First, however, we consider other astrophysical sources which do not produce antineutrino fluxes comparable with the terrestrial flux.

Antineutrinos are absent from all major solar reactions. This is important because even if their production rate was suppressed by 10^{-4} – 10^{-5} relative to solar neutrinos, their flux on Earth could equal that of terrestrial antineutrinos.

In describing this suppression, we note that all 18 reactions in the p–p cycle, and the CNO cycle involve the production of neutrinos or photons only²⁸; thus antineutrino production is divorced from the primary energy production mechanisms of the Sun, and we must consider other weak interaction processes involving neutrino pair production²⁸, however, these are strongly suppressed, and are not significant until temperatures of order $\sim 10^9 \text{ K}$ (ref. 28). At solar temperatures they are down by factors of the order of $\leq 10^{-10}$. Because of its distance from Earth, in order for fluxes due to β^- decays in the Sun to compete with fluxes for terrestrial β^- decay, radioactive isotopic abundances on the Sun must be greater by a factor of $\sim 10^6$. No solar model predicts such abundances.

Having demonstrated that essentially no solar antineutrino flux exists, we may consider other point sources. As the terrestrial antineutrino flux is of the order of 10^{-4} times the solar neutrino flux, the antineutrino luminosity of any source must be of the

order of $L \sim 10^{-4} (L_{\text{sun}})(R/R_{\text{sun}})^2$ to compete with the terrestrial antineutrino flux. This rules out all known continuous point sources as significant sources of antineutrinos. Even the flux of neutrino and antineutrinos from the galactic centre is, for energies of interest, expected to be orders of magnitude smaller than the atmospheric neutrino flux which is much smaller than the terrestrial flux²².

Only supernovae explosions, which emit $\sim 10^{53}$ erg of energy, mostly in the form of neutrinos and antineutrinos with energies ~ 10 MeV (ref. 29), are, in principle, energetic enough to produce fluxes comparable or larger than the solar or terrestrial fluxes. However, calculations using existing supernovae neutrino detectors demonstrate that a 1 kton detector will only be sensitive to supernovae in our Galaxy²⁹. These explosions are expected to occur at a rate of only 1 every 10–30 yr, making individual detection problematic.

If supernovae have been occurring at a constant rate of 1 per 15 yr per galaxy for most of the history of the Universe (say $\sim 10^{10}$ yr), their cumulative effects would have produced a diffuse background of neutrinos and antineutrinos. The neutrino background is hidden by the high-energy portion of the solar neutrino flux. However, the antineutrino background is, in principle, separable from the terrestrial flux.

We can calculate the present antineutrino luminosity at the Earth, taking into account the effects of Universe expansion as follows³⁰. Each particle emitted with energy E at cosmic time t_i will be redshifted at observation time to the energy $ER(t_i)/R(t_0)$, where $R(t)$ represents the cosmic scale factor describing isotropic homogeneous expansion. Similarly, antineutrinos emitted at intervals δt_i will arrive at time intervals $\delta t_i R(t_0)/R(t_i)$. Hence the apparent luminosity (energy/unit area/time) of a supernova at co-moving coordinate r_i (relative to the Earth) is

$$I = \frac{LR^2(t_i)}{4\pi R^4(t_0)r_i^2} \quad (9)$$

where L is the absolute luminosity of the supernovae in antineutrinos.

Now the number of supernovae observed at t_0 , which occurred between time $t_i - \delta t_i$ and t_i , is:

$$dN = n(t_i)4\pi R^2(t_i)r_i^2 \delta t_i \quad (10)$$

where n is the number density of supernovae per unit time.

Thus the total energy density of neutrinos and antineutrinos is measured by an observer on Earth at time t_0 is:

$$\rho_{\nu, \bar{\nu}} = \int_{t_s}^{t_0} n(t_i) L \left[\frac{R(t_i)}{R(t_0)} \right]^4 dt_i \quad (11)$$

where t_s is the initial time of supernovae occurrence, which can be taken to zero without affecting the numerical result. If supernova frequency has been constant, so that $n(t_i) = n(t_0)[R(t_0)/R(t_i)]^3$, equation (11) can be integrated to yield:

$$\rho_{\nu, \bar{\nu}} = \eta(t_0) L \int_0^{t_0} \frac{R(t)}{R(t_0)} dt = \frac{3}{5} \eta(t_0) L t_0 \quad (12)$$

Thus the net effect of expansion is to reduce the neutrino energy density by a factor 3/5 compared with what could be expected in the absence of expansion. Now assuming $\sim 10^{53}$ erg emitted per supernova, and that $\eta(t_0) \sim 10^{-3} \text{ Mpc}^{-3} \text{ yr}^{-1}$ (this is only a rough estimate, the actual densities of galaxies may be less. The frequency we assume may also be too low³¹), equation (12) gives

$$\rho_{\nu, \bar{\nu}} = 3 \times 10^{-14} \text{ erg cm}^{-3} = 2 \times 10^{-8} \text{ MeV cm}^{-3} \quad (13)$$

The mean energy per antineutrino (neutrino) is ~ 10 MeV at the time of emission is now redshifted by the same factor of 3/5 of equation (12). Hence, the total flux density of neutrinos and antineutrinos through a small volume at the Earth's surface is now

$$N(\nu, \bar{\nu}) = \frac{\rho_{\nu, \bar{\nu}}}{6} \sim 10^2 \text{ cm}^{-2} \text{ s}^{-1} \quad (14)$$

If roughly half the energy is emitted in the form of antineutrinos²⁹ then the number density of antineutrinos at the Earth's surface at present is $\sim 50 \text{ cm}^{-2} \text{ s}^{-1}$, with a spectrum peaked at ~ 6 MeV. This number could increase if the frequency of supernovae increased at earlier times.

This flux is not large enough to warrant an experiment for its determination. However, antineutrino detectors may eventually be sensitive to these antineutrinos. As their mean energy is significantly larger than terrestrial antineutrinos, they may yield an effect at the $\sim 1\%$ level for inverse β -decay type detectors. If this small effect could be measured, it would yield the first estimate of the time-averaged supernovae frequency per galaxy over the lifetime of the Universe.

Conclusions

The neutrino and antineutrino fluxes described here are shown on Fig. 2. Neutrinos provide information on the Sun and stellar processes, and antineutrinos provide information on the Earth, and perhaps on supernova abundance. Both processes provide information which is probably inaccessible by means other than weak interaction probes. The absence of electromagnetic interactions for neutrinos and antineutrinos implies that they may open for investigation an orthogonal and complementary component of the Universe to that available through electromagnetic detection. Moreover, neutrinos and antineutrinos may account for much of the energy density of the Universe, and hence may have a vital role in its dynamics.

Yet the aspects of the weak interactions that make them so useful for investigating the interior of the Sun and Earth also hinder their measurement. Detection of terrestrial antineutrinos will require sophisticated new technology. We can envision several generations of experiments aimed at measuring different parts of the antineutrino spectrum. If this proves possible, we may obtain the first fundamental empirical data on the Earth's interior structure and dynamics, including measurements of heat production, convection in the upper mantle, geological data on interior composition and perhaps information on local concentrations of, or global anisotropies in radioactive abundances. Also, it may be possible to gain information on the abundance and frequency of supernovae, on the antineutrino component of cosmic rays on possible neutrino oscillations, and to verify theoretical predictions of low-energy weak interaction cross-sections.

It is a long step from even a single measurement to the programme outlined above. Indeed, even rudimentary weak interaction astronomy and geophysics may not become practicable for some time. In principal however, neutrino and antineutrino astronomy and geophysics can open vast new windows for exploration above us and below.

We thank our colleagues, including M. Turner, S. Bludman, F. Wilczek, E. Fireman, R. Clayton, K. Lande, R. Davis, A. Yahil, R. Wagoner, E. Purcell and N. Isgur, for useful discussions. In particular, M. Turner corrected an error in initial flux calculations and R. Davis informed us on experimental matters and of past work on this subject. L.M.K. and D.S. also thank the Aspen Institute of Physics where this work was initiated. Finally, we thank W. Hampel for criticism which improved the paper in many ways—in particular by pointing out the backgrounds due to stopped muons.

1. Bahcall, J. N. *Rev. mod. Phys.* **50**, 881–903 (1978).
2. Davis, R. Jr, Evans, J. C. & Cleveland, B. T. in *Conf. Proc. Neutrinos-78* (ed. Fowler, E. C.) (Purdue University Press, 1978).
3. Eders, G. *Nucl. Phys.* **78**, 657–662 (1966).
4. Marx, G. *Czech. J. Phys.* **B19**, 1471–1479 (1969).
5. Marx, G. & Lux, I. *Acta phys. acad. sci. hung.* **28**, 63–70 (1970).
6. Avilez, C., Marx, G. & Fuentes, B. *Phys. Rev. D* **23**, 1116–1117 (1981).
7. *Conf. Proc. Cosmological and Geological Implications of Isotope Ratio Variations* (National Academy of Sciences Publ. 572, 1958).
8. Jeffreys, H. *The Earth* 5th edn (Cambridge University Press, 1970).
9. Mackenzie, D. P. *Scient. Am.* **249**, 67–78 (1983).
10. Sclater, J. G., Parsons, B. & Jaupart, C. *J. geophys. Res.* **86**, 535–552 (1981).
11. *Handbook of Chemistry & Physics* (Chemical Rubber Co., Florida, 1982).
12. Jaupart, G., Sclater, J. G. & Simmons, G. *Earth planet. Sci. Lett.* **52**, 328–344 (1981).

13. Sclater, S. G., Jaupart, C. & Galson, D. *Rev. Geophys. Space Sci.* **18**, 269–311 (1980).
14. Lubimova, E. B. in *Earth's Crust and Upper Mantle* (ed. Hart, P. J.) (Washington, DC, 1969).
15. Mikaelyan, L. A., Tsinoev, B. G. & Bereyay, A. A. *Soviet J. nucl. Phys.* **6**, 254–256 (1968).
16. Segre, E. *Nuclei and Particles* (Benjamin, Reading).
17. Particle Data Group *Phys. Lett.* **111B** (1982).
18. Fireman, E. L. in *Proc. int. Cosmic Ray Conf. Kyoto*, Vol. 13 (1979).
19. Frampton, P., Glashow, S. L. & Yildiz, A. (eds) *First Workshop on Grand Unification* (Mathematical Science Press, Brookline, 1980).
20. Cherry, M. L. *et al. J. Phys. G, nucl. Phys.* **8**, 879–885 (1982).
21. Charalambus, S. *Nucl. Phys.* **A166**, 145–161 (1971).
22. Margolis, S. H., Schramm, D. N. & Silverberg, R. *Astrophys. J.* **221**, 990–1002 (1978).
23. Stecker, F. W. *Astrophys. J.* **228**, 919–927 (1979).
24. Gaisser, T. K., Stanev, T., Bludman, S. & Lee, H. Preprint (Bartol Research Inst., 1983).
25. Bionta, R. M. *et al. Phys. Rev. Lett.* **51**, 27–30 (1983).
26. Laney, C. A. *Metallic and Industrial Mineral Deposits* (McGraw-Hill, New York, 1966).
27. Skinner, B. J. *Earth Resources* (Prentice-Hall, New Jersey, 1976).
28. Reeves, H. in *Stars and Stellar Systems* (eds Aller, L. H. & McLaughlin, D. P.) (University of Chicago Press, 1965).
29. Bethe, H. A., Yahil, A. & Brown, G. E. *Astrophys. J. Lett.* **262**, L7–L10 (1982).
30. Weinberg, S. *Gravitation and Cosmology* (Wiley, New York, 1972).
31. Tammann, G. in *Supernovae* (ed. Schramm, D. N.) (Reidel, Dordrecht, 1977).

ARTICLES

Origin of diamonds in old enriched mantle

S. H. Richardson*, J. J. Gurney†, A. J. Erlank† & J. W. Harris‡

* Center for Geoalchemy, Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

† Department of Geochemistry, University of Cape Town, Rondebosch 7700, South Africa

‡ Department of Applied Geology, University of Strathclyde, Glasgow G1 1XJ, UK

Sub-calcic garnets encapsulated by diamonds from relatively young (90 Myr) kimberlites in southern Africa, yield ancient Sm–Nd and Rb–Sr model ages (3,200–3,300 Myr). The chemistry and distribution of these and associated sub-calcic garnets from kimberlite concentrate, indicate diamonds formed following enrichment of residual sub-cratonic mantle such as that remaining after widespread extraction of 3,500-Myr komatiitic lavas.

ALTHOUGH diamonds have been recovered directly from kimberlite pipes for more than a century, their relation to host kimberlite has remained controversial^{1–3}. The debate centres on whether diamonds are phenocrysts or xenocrysts in the proto-kimberlite magma. More broadly, diamond crystallization may be temporally and spatially associated with kimberlite genesis, or, alternatively, segments of mantle within the diamond stability field may be littered with diamonds, so that sampling by kimberlite is strictly accidental. Diamond is, in fact, only a trace constituent of kimberlite, with abundances varying from zero to an order-of-magnitude maximum of 1 p.p.m. Moreover, in southern Africa, kimberlites erupted within the boundaries of the Archaean craton are diamondiferous, while those in adjacent younger mobile belts are barren⁴.

A small fraction of diamonds (<1%) contains syngenetic monomineralic and rarely multiminerallitic inclusions belonging to peridotitic or eclogitic parageneses. The peridotitic inclusion assemblage, which includes olivine, orthopyroxene, chromite and distinctive lilac chrome-pyroxene garnet, is on average far more abundant than the eclogitic assemblage of orange pyrope-almandine garnet and omphacitic clinopyroxene^{5–7}. This categorization and relative abundance are apparently also reflected in diamond carbon isotope compositions⁸, most of which have been recorded in inclusion-free diamonds⁹. Thus, constraints on diamond origin derived from such inclusions may be considered applicable to most inclusion-free diamonds.

Diamond inclusion assemblages parallel the mineralogy of the two most important categories of mantle xenolith in kimberlites, garnet peridotite and eclogite. However, there are significant differences in chemical composition between the dominant peridotitic diamond inclusions and constituent minerals of garnet peridotites. For example, garnets in diamond have significantly higher Mg and Cr, and lower Ca, contents than their peridotite xenolith counterparts, which are generally saturated with Ca (Fig. 1). The former are here referred to as sub-calcic garnets. Lilac garnets of similar but less extreme composition have also been found in kimberlite concentrate¹⁰. The distribution of these discrete sub-calcic garnets in southern African kimberlites generally mirrors that of diamonds, implying an intimate genetic association¹¹.

Previous radiogenic isotope studies on trace constituents of diamond-related materials have been viewed with some scepticism, given identification and potentially severe contamination problems. For example, a U–Pb study of presumed sulphide inclusions in diamonds yielded model Pb ages in excess of 2,000 Myr but was clouded by uncertainty regarding inclusion identity and paragenesis¹². Similarly, results of a He isotopic study of industrial-grade diamonds required speculation on He entrapment by diamonds soon after accretion of the Earth¹³.

We have investigated the Sm–Nd and Rb–Sr isotopic systematics of carefully screened sub-calcic garnets in concentrate and diamonds from two of the best documented localities in southern Africa, Kimberley (28°46'S, 24°48'E) and Finsch (28°16'S, 23°06'E; ~160 km WNW of Kimberley). These kimberlites are respectively representative of two classic types of kimberlite (basaltic and micaceous) recognized on the basis of petrographic, chemical and isotopic characteristics^{14,15}. Our consistent results preclude a genetic connection between diamonds and host kimberlites and provide strong age and chemical constraints on an ultimate origin of diamonds in old, enriched mantle.

Sub-calcic garnets in diamonds

Garnet inclusions in diamond are easily recognized *in situ* by their distinctive colours. Peridotitic inclusions, of which 10–20% are garnets, comprise more than 98% of inclusions in Finsch diamonds⁷ and more than 90% of those in Kimberley diamonds¹⁶. A comprehensive study of inclusion major element compositions¹⁷, revealed that more than 95% of peridotitic garnets in Finsch diamonds have extremely high Mg and Cr and low Ca contents. Associated olivine and orthopyroxene inclusion compositions are also extremely magnesian with average values of Mg/(Mg+Fe) of 0.941 and 0.949 respectively¹⁷. Peridotitic diopside inclusions are conspicuous by their absence, emphasizing the residual major element character of the peridotitic inclusion paragenesis^{5–7}.

Lilac garnet, inclusion bearing, diamonds were selected from the –6 +5 and –7 +6 diamond sieve fractions¹⁸ of the general production from Finsch and the Kimberley Pool. The latter represents output from Bultfontein, De Beers, Dutoitspan and Wesselton, the four kimberlite pipes currently mined at Kimberley, which are close enough at surface to have possibly had the

* Present address: Institut de Physique du Globe, Université de Paris VI, 4 Place Jussieu, 75230 Paris Cedex 05, France.

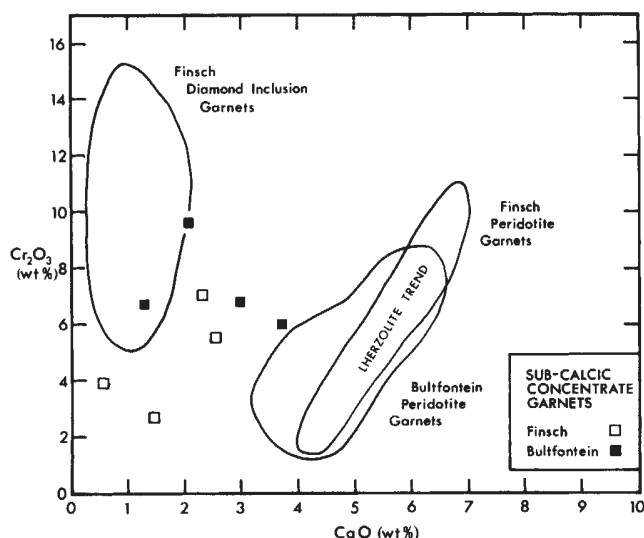


Fig. 1 Distribution of CaO and Cr₂O₃ in selected sub-calcic garnets from Bultfontein (Kimberley Pool) and Finsch kimberlite heavy mineral concentrates. Also shown for comparison are fields enclosing greater than 95% of sub-calcic garnet inclusions in Finsch diamonds¹⁷, garnets in Finsch peridotites²⁰ and garnets in Bultfontein peridotites³⁵ (coarse and deformed lherzolites and harzburgites). Positive correlations described by peridotite garnets have become known as the lherzolite trend³⁶ as they reflect saturation with a diopside component, regardless of the modal abundance of diopside. Sub-calcic concentrate garnets scatter away from this trend towards diamond inclusion garnets, suggesting diopside-free host assemblages.

same in-depth source¹⁶. Garnets were liberated by breaking diamonds in a steel cracker, and recovered intact, or occasionally in a few pieces, under a binocular microscope. Whole inclusions, often displaying diamond controlled morphologies, averaged 100–200 μ m in diameter and 10–15 μ g in weight. In view of small inclusion size, coupled with Nd and Sr concentrations of only a few p.p.m., as determined in a pilot study¹⁹, several hundred garnets from each locality were composited for isotope analysis.

Inclusions from otherwise internally flawless diamonds were found to be in pristine condition. Partial alteration of some inclusions in other diamonds could generally be traced to emplacement-related fractures in diamonds, which were open to penetration by deuteric solutions. Approximately 15% of Finsch and 30% of Kimberley inclusion garnets were rejected on the basis of the slightest trace of alteration, discoloration, coating or surface pitting visible during microscopic examination of all grain surfaces before and after successive ultrasonic cleaning in ultrapure 2.5 M HCl, 5% HF, 2.5 M HCl, water and acetone. For the benefit of a duplicate determination, Finsch garnets were further divided into two fractions (Table 1) before final cleaning.

Sub-calcic garnets from kimberlite

Red to lilac chrome-pyropite garnet crystals up to a few millimetres in diameter are generally abundant in the heavy mineral concentrate recovered from kimberlites. Most of these fall in the range of compositions for peridotite garnets, which display a positive correlation between Ca and Cr (Fig. 1). However, a significant Hfraction of lilac garnets scatter towards and partially overlap most diamond inclusion compositions^{10,11}. At Finsch, such sub-calcic garnets comprise ~100 p.p.m. of the kimberlite as a whole¹⁰, as compared with a diamond abundance of 0.2 p.p.m. (ref. 20).

For this study, several sub-calcic garnets from Bultfontein (Kimberley Pool) and Finsch concentrates, were identified on the basis of full electron microprobe analysis of fragments split from pre-selected lilac garnet grains. Ca and Cr contents of

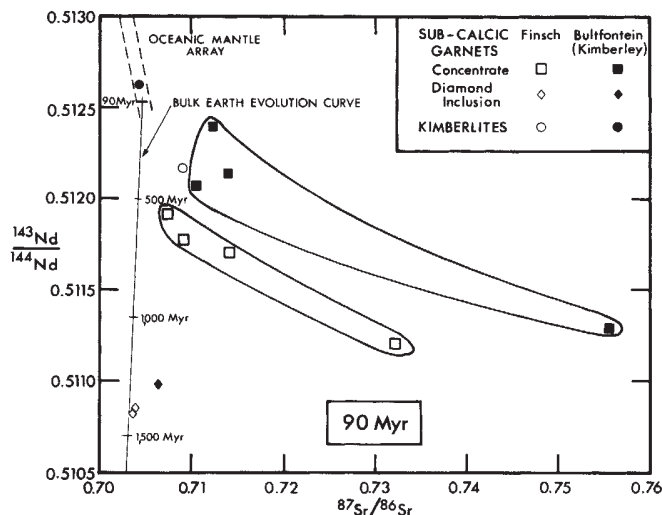


Fig. 2 Nd-Sr isotope correlation diagram, constructed at 90 Myr (approximate time of kimberlite emplacement)²³, for sub-calcic garnets. Concentrate garnets are single grains from Bultfontein (Kimberley Pool) and Finsch kimberlite heavy mineral concentrates, while inclusion garnets are composites of many specimens from Kimberley Pool and Finsch diamonds. The oceanic mantle array^{37,38}, a bulk Earth evolution curve, based on average chondritic Nd isotopic evolution³⁹, and host kimberlite values¹⁹ are shown for reference. Diamond inclusion garnets clearly bear no relation to host kimberlites, or any other modern mantle-derived volcanics. Model age arguments (Figs 3 and 4) indicate ancient diamond formation in enriched mantle 3,200–3,300 Myr ago. Concentrate garnets show crude inverse arrays, extending from extreme high ⁸⁷Sr/⁸⁶Sr and low ¹⁴³Nd/¹⁴⁴Nd ratios back towards peridotite garnet¹⁹, kimberlite and oceanic mantle values. Values for micaceous kimberlites from southern Africa¹⁵, as well as kimberlites and lamproites from western Australia³¹, plot in the vicinity of the top left end of these arrays, as for the Finsch kimberlite.

individual grains used for isotope analysis are shown in Fig. 1. While grains have interiors that are extremely fresh and uncracked, they invariably carry remnants of kelyphite crusts or underlying pitted surfaces. All such exterior surfaces were removed using thin sliver splitting techniques under a binocular microscope. Remaining whole 2–4 mg grains were then subjected to successive ultrasonic cleaning in ultrapure 2.5 M HCl, 5% HF, 2.5 M HCl, water and acetone, followed by piece-by-piece grain size reduction in a steel cracker and final inspection and cleaning.

Analytical techniques

Chemical and mass spectrometric techniques for determination of concentrations and isotope ratios are as described in refs 21 and 22 but with provision for precise determination of Sm/Nd and ¹⁴³Nd/¹⁴⁴Nd ratios on totally spiked samples¹⁹. Apart from pilot study extremes, available quantities of garnet Sr and Nd were 2–35 ng and 10–45 ng respectively. In-run precision ($2\sigma_{\text{mean}}$) of 0.007% for ⁸⁷Sr/⁸⁶Sr and 0.005% for ¹⁴³Nd/¹⁴⁴Nd was attainable for as little as 10 ng of each element from the same garnet sample in runs with standard ⁸⁸Sr and ¹⁴⁴Nd ion beam intensities of 1×10^{-11} A and 3×10^{-12} A respectively, lasting up to 2 h. For garnet Sr contents in the 2–4 ng range, ⁸⁸Sr ion beam intensities were reduced to 3×10^{-12} A with corresponding decrease in precision. No obvious differential instrumental bias attributable to use of this scaled reduction in intensity, has been detected within the precision of standard runs. Total procedural blank corrections (see Table 1) are generally barely significant other than for Rb. Alkali contents of small garnet samples are only 0–2 times minimum blank levels, resulting in large relative but trivial absolute uncertainties in Rb concentrations which are inherently close to zero.

Table 1 Rb–Sr and Sm–Nd concentrations* and isotope ratios† for sub-calcic garnets in heavy mineral concentrates and diamonds from Kimberley Pool‡ and Finsch kimberlite pipes

	wt(mg)	Rb	Sr	⁸⁷ Rb/ ⁸⁶ Sr	⁸⁷ Sr/ ⁸⁶ Sr _p	⁸⁷ Sr/ ⁸⁶ Sr _i	Sm	Nd	¹⁴⁷ Sm/ ¹⁴⁴ Nd	¹⁴³ Nd/ ¹⁴⁴ Nd _p	¹⁴³ Nd/ ¹⁴⁴ Nd _i
Concentrate garnets											
Bultfontein											
Composite§	13.25	0.022	2.97	0.021	0.71397 ± 4	0.71394	2.797	9.042	0.1870	0.512253 ± 15	0.512143
JJG/SR4	2.67	0.006	1.23	0.014	0.71020 ± 46	0.71018	1.982	8.202	0.1462	0.512202 ± 20	0.512116
JJG/SR10	3.73	0.007	0.671	0.030	0.71230 ± 23	0.71226	1.558	3.111	0.3027	0.512574 ± 25	0.512396
JJG/SR20	2.77	0.001	0.993	0.003	0.75510 ± 33	0.75510	1.294	3.871	0.2020	0.511409 ± 35	0.511290
Finsch											
FINCON5	3.27	0.001	0.642	0.003	0.70734 ± 20	0.70734	0.8288	3.094	0.1619	0.512015 ± 37	0.511920
FINCON7	3.40	0.002	2.66	0.002	0.70889 ± 24	0.70889	0.7954	7.327	0.06561	0.511800 ± 17	0.511761
FINCON11	2.26	0.004	5.44	0.002	0.73213 ± 7	0.73213	1.714	12.149	0.08526	0.511250 ± 17	0.511200
FINCON/07/18	2.90	0.002	12.17	0.0004	0.71424 ± 3	0.71424	3.342	15.282	0.1322	0.511781 ± 16	0.511703
Diamond inclusion garnets											
Kimberley Pool‡											
Composite	5.93	0.003	1.985	0.004	0.70621 ± 5	0.70621	0.6115	2.920	0.1265	0.511058 ± 36	0.510984
Finsch											
Composite P¶	3.81	0.013	6.379	0.006	0.70380 ± 4	0.70379	1.287	6.569	0.1184	0.510919 ± 15	0.510849
Composite C#	3.79	0.015	8.610	0.005	0.70356 ± 8	0.70355	1.186	6.301	0.1137	0.510892 ± 22	0.510825

* Concentrations in p.p.m. (µg per g) after correction for minimum applicable blanks [Bultfontein concentrate garnets: Rb(26 pg), Sr(100 pg), Sm(1.5 pg), Nd(2.3 pg); Finsch concentrate garnets and all diamond inclusion garnets: Rb(20 pg), Sr(15 pg), Sm(1.5 pg), Nd(2.3 pg)]. Estimated analytical uncertainties in concentrations (ignoring small-sample weighing errors) indicated by number of significant digits.

† ⁸⁷Sr/⁸⁶Sr ratios normalized to 0.70800 for E & A SrCO₃ [average measured value = 0.70783 ± 3 (2σ, N = 5, 1981–83) using ⁸⁶Sr/⁸⁸Sr = 0.1194] and corrected for an identified reagent blank ⁸⁷Sr/⁸⁶Sr of 0.70783 ± 15. Latter corrections all less than in-run precision (2σ_{mean} errors corresponding to least significant digits) except for JJG/SR20. ¹⁴³Nd/¹⁴⁴Nd ratios normalized to 0.51264 for BCR-1 [average measured value = 0.512622 ± 14 (2σ, N = 6, spiked, 1982–83) using ¹⁴⁶Nd/¹⁴⁴Nd = 0.7219]. Blank effects on Nd isotope ratios insignificant. Initial ratios calculated for an approximate kimberlite emplacement age of 90 Myr (ref. 23) using λ_{Rb} = 1.42 × 10^{−11} yr^{−1} and λ_{Sm} = 6.54 × 10^{−12} yr^{−1}. Differential effects of a slightly older age for Finsch (ref. 15) are negligible.

‡ Kimberley Pool includes Bultfontein, De Beers, Dutoitspan and Wesselton pipes. Mined kimberlite is channelled through common diamond recovery plant.

§ Pilot study composite of six 1–3 mg sub-calcic garnet grains from concentrate.

|| Composite of ~600 flawless as well as irregular and fragmented sub-calcic garnet inclusions (average weight = 10 µg). Garnet inclusions enclosing minute chromite grains excluded.

¶ Composite of ~200 flawless sub-calcic garnet inclusions (average weight = 19 µg).

Composite of ~300 irregular and fragmented sub-calcic garnet inclusions (average weight = 13 µg) of which ~10% enclosed minute chromite grains (residual after garnet dissolution with total weight <10 µg).

Sm–Nd and Rb–Sr isotopic systematics

Sub-calcic garnet ¹⁴³Nd/¹⁴⁴Nd and ⁸⁷Sr/⁸⁶Sr ratios (Table 1) are illustrated on a Nd–Sr isotope correlation diagram (Fig. 2) constructed at 90 Myr, the approximate time of host kimberlite emplacement²³. Corresponding age corrections are small for ¹⁴³Nd/¹⁴⁴Nd, since Sm/Nd ratios are low (0.1–0.5), and generally insignificant for ⁸⁷Sr/⁸⁶Sr, since Rb/Sr ratios are negligible (<0.01).

The immediate first-order observation is that diamond inclusion garnets, concentrate garnets and host kimberlites have distinct isotopic signatures. Sub-calcic diamond inclusion garnets have by far the lowest ¹⁴³Nd/¹⁴⁴Nd ratios measured in any modern mantle-derived materials. Such inclusion garnets clearly carry an ancient Nd isotopic component and syngenetic host diamonds must necessarily be xenocrysts in kimberlites. Sub-calcic concentrate garnets have a large spread in ¹⁴³Nd/¹⁴⁴Nd ratios negatively correlated with extremely high ⁸⁷Sr/⁸⁶Sr ratios. Such concentrate garnets are also necessarily xenocrysts in kimberlites but are clearly isotopically differentiated from diamond inclusion garnets. Yet the two types of sub-calcic garnet show sympathetic differences in isotopic signature between localities. Specifically, the Finsch concentrate garnet array and diamond inclusion composites lie at relatively lower ¹⁴³Nd/¹⁴⁴Nd and ⁸⁷Sr/⁸⁶Sr ratios than their Kimberley counterparts.

The next obvious observation is that both diamond inclusion and concentrate garnet ⁸⁷Sr/⁸⁶Sr ratios are entirely unsupported by uniformly low intrinsic Rb contents, even over the age of the Earth. Therefore, the garnets must have inherited such ⁸⁷Sr/⁸⁶Sr ratios by crystallization from and/or subsequent diffusive equilibration with mantle hosts which had higher time-integrated Rb/Sr ratios for an appropriate period of time. As Sm, Nd, Rb and Sr diffusion rates through diamond are assumed to be negligible, diamond crystallization effectively isolates inclusions from further interaction with the host. Therefore, the difference in isotopic signature between the two types of sub-calcic garnet could simply be a function of exposure time in a common host chemical environment, which is similar but not identical from place to place.

Overall, sub-calcic garnets are endowed with trace element abundances which belie their residual major element compositions. Concentrations of Nd (3–15 p.p.m.) and Sr (0.6–12 p.p.m.) range up to significantly higher values than those for peridotite xenolith garnets (1–4 p.p.m. Nd, 0.5–1 p.p.m. Sr; ref. 19). Moreover, relatively low garnet Sm/Nd ratios suggest strongly light rare earth element (LREE) enriched patterns which are the opposite of those observed in peridotite xenolith garnets²⁴. On a Sm–Nd isochron diagram (not shown), diamond inclusion garnets show restricted variation in low ¹⁴⁷Sm/¹⁴⁴Nd and ¹⁴³Nd/¹⁴⁴Nd ratios. In contrast, concentrate garnets show a substantial range of ¹⁴³Nd/¹⁴⁴Nd ratios and great scatter.

Representation of garnet Rb–Sr data on a Rb–Sr isochron diagram is not particularly useful, since Rb/Sr ratios are all extremely low (0.0001–0.01) and absolute values are imprecise because of uncertainties in vanishingly small Rb contents. Nevertheless, diamond inclusion garnets show an apparent positive correlation between ⁸⁷Sr/⁸⁶Sr ratio (0.7036–0.7062) and reciprocal Sr concentration (not shown), and by inference precursor host Rb/Sr ratio. In contrast, concentrate garnets show an extreme range of ⁸⁷Sr/⁸⁶Sr ratios and no coherent relationship with Sr concentrations.

By all accounts, Sm–Nd and Rb–Sr concentrations in diamond inclusion garnets show consistent relationships with isotope ratios, whereas in concentrate garnets they do not. Therefore, only syngenetic inclusion garnet Sm/Nd ratios may be used to derive Sm–Nd model ages for diamond formation. Time-integrated precursor Rb–Sr ratios may in turn be used to calculate differential Rb–Sr model ages for precursor formation.

Age constraints on diamond formation

Sm–Nd model age relationships are best illustrated in a ¹⁴³Nd/¹⁴⁴Nd evolution diagram (Fig. 3). Intersections of garnet and reference curves yield Sm–Nd model ages for syngenetic garnet and diamond derivation from the hypothetical bulk Earth reservoir. Ages thus obtained are 3,410 ± 80 Myr for the Kimberley diamond inclusion composite and 3,320 ± 30 and 3,190 ± 50 Myr for the two Finsch diamond inclusion composites (P

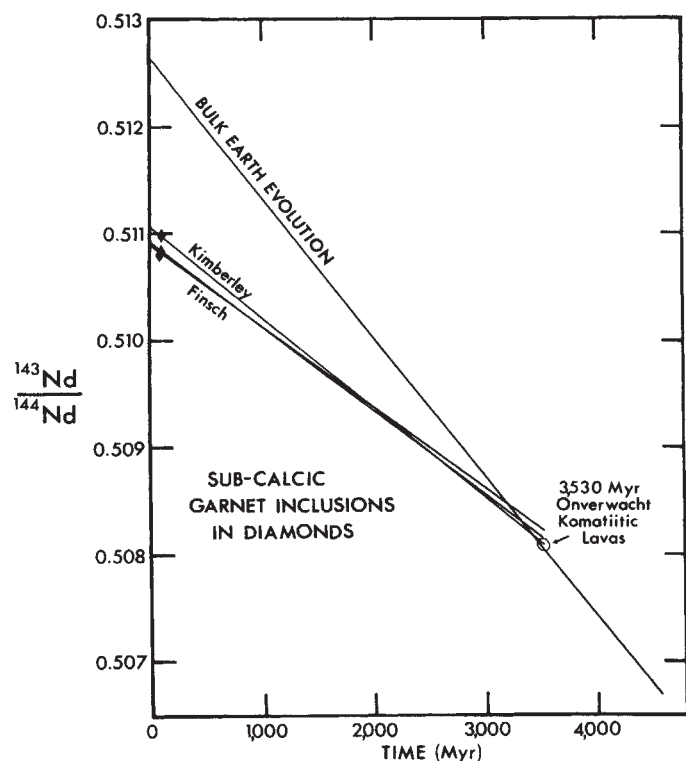


Fig. 3 $^{143}\text{Nd}/^{144}\text{Nd}$ evolution diagram showing evolution curves for sub-calcic garnets in Kimberley Pool and Finsch diamonds. Curves are constructed from measured $^{143}\text{Nd}/^{144}\text{Nd}$ and $^{147}\text{Sm}/^{144}\text{Nd}$ ratios, using $\lambda_{\text{Sm}} = 6.54 \times 10^{-12} \text{ yr}^{-1}$, and terminated for convenience at 3,530 Myr. Diamond symbols at 90 Myr mark the approximate time of diamond transport to surface by kimberlite²³. The bulk Earth reference curve is constructed using appropriate present-day chondritic values ($^{143}\text{Nd}/^{144}\text{Nd} = 0.512638$, $^{147}\text{Sm}/^{144}\text{Nd} = 0.1966$)^{39,40,26}. The Onverwacht komatiite age (3,530 Myr) and initial $^{143}\text{Nd}/^{144}\text{Nd}$ ratio are from refs 25 and 26 using an error ellipse representation⁴¹. Intersections of garnet and bulk Earth curves give Sm–Nd model ages for garnet encapsulation by diamond. Using a simple combination of analytical and model uncertainties, a conservative overall estimate of diamond formation age is $3,300 \pm 200$ Myr. Temporal, spatial and major element compositional relationships suggest diamond formation occurred subsequent to komatiite generation in Archaean mantle beneath the Kaapvaal craton. Trace element and Sr isotopic relationships (Fig. 4) require introduction of an enriched component (high Rb/Sr and low Sm/Nd ratios) following komatiite generation as a precursor to diamond formation.

and C). Analytical uncertainties are as propagated from those for both $^{143}\text{Nd}/^{144}\text{Nd}$ ($2\sigma_{\text{mean}}$) and $^{147}\text{Sm}/^{144}\text{Nd}$ ($<0.1\%$) measurements but effectively dominated by the former. In principle, results for the two Finsch composites were meant to represent a first-order duplicate determination, and in this sense the ages obtained are very similar, if not exactly within nominal errors. In practice, the second composite (C) consisted of irregular and fragmented garnet inclusions and included a minor proportion enclosing minute chromite grains, residual after dissolution (Table 1). The marginally younger age could be due to sample character, imperfect screening, minor contamination or simply underestimated analytical errors.

Further systematic model age uncertainties arise from the assumption of a bulk Earth reference curve, since neither an exactly chondritic initial Earth nor a subsequent undifferentiated mantle source for garnet and diamond can be presumed. The Nd isotopic character of real temporally and spatially associated mantle may usefully be represented by that of 3,530-Myr old Onverwacht Group volcanics^{25,26}, from the Barberton Mountain Land in southern Africa. These komatiitic volcanics are the best preserved remnant of such magmas which appear to have been emplaced across the entire Kaapvaal craton²⁷. For a mantle evolution curve passing through the Onverwacht datum (Fig.

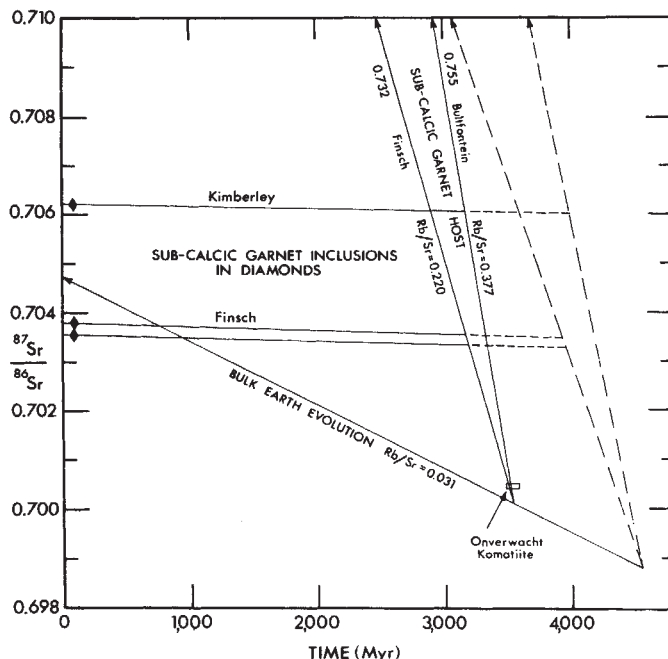


Fig. 4 $^{87}\text{Sr}/^{86}\text{Sr}$ evolution diagram showing evolution curves for sub-calcic garnets in Kimberley Pool and Finsch diamonds. Curves are constructed from measured $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{87}\text{Rb}/^{86}\text{Sr}$ ratios, using $\lambda_{\text{Rb}} = 1.42 \times 10^{-11} \text{ yr}^{-1}$. Intrinsic garnet Rb/Sr ratios are so low that corresponding evolution curves are virtually flat. Diamond symbols as in Fig. 3. The bulk Earth reference curve is drawn using a chondritic meteorite initial $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of 0.69880 and a present-day silicate Earth $^{87}\text{Sr}/^{86}\text{Sr}$ ratio of 0.70475 derived from the mantle Nd–Sr correlation (see Fig. 2 and ref. 42). The Onverwacht datum is a combination of initial $^{87}\text{Sr}/^{86}\text{Sr}$ (from ref. 43) and 3,530 Myr Sm–Nd isochron age (from refs 25 and 26). Hypothetical sub-calcic garnet host (diamond precursor) evolution curves are drawn between the bulk Earth curve and extreme $^{87}\text{Sr}/^{86}\text{Sr}$ ratios measured in sub-calcic garnets from kimberlite heavy mineral concentrates. Solid curves are drawn between the bulk Earth at 3,530 Myr and the highest concentrate garnet $^{87}\text{Sr}/^{86}\text{Sr}$ ratios (at 90 Myr) for Bultfontein (0.755) and Finsch (0.732). Corresponding time-integrated Rb/Sr ratios are indicated. Dashed curves emanating from the bulk Earth at 4,550 Myr are shown for reference. Intersections of respective diamond and solid sub-calcic garnet host curves give a common Rb–Sr model age for garnet encapsulation by diamond of 3,200 Myr, predicated on an approximate 300 Myr time differential between precursor enrichment (following komatiite generation) and diamond crystallization.

3) and either slightly steeper (depleted mantle with higher Sm/Nd) or shallower (enriched mantle with lower Sm/Nd) than the reference curve, garnet model ages would be respectively increased or decreased by ~ 100 Myr on average. In effect, the ages of Kimberley and Finsch garnets are not resolved within reasonable errors and approximately synchronous diamond formation is suggested. With simple combination of analytical and model uncertainties, a conservative overall estimate of sub-calcic, garnet-bearing diamond formation age, is $3,300 \pm 200$ Myr.

The association between diamond and komatiite genesis¹¹ is not only temporal and spatial, but also compositional. For example, diamond inclusion olivine Mg/(Mg + Fe) ratios are essentially identical to those for liquidus olivines in komatiites²⁹. While the mineralogy of residual mantle or high-pressure cumulates remaining after komatiite generation is controversial²⁸, highly magnesian major element compositions are not disputed. Nd/Sm and Rb/Sr ratios equal to or less than bulk Earth values may also be inferred for such depleted residual mantle. Such trace element characteristics are the opposite of those inferred from measured sub-calcic garnet Nd/Sm and $^{87}\text{Sr}/^{86}\text{Sr}$ ratios. Therefore, enrichment of such residues is required after komatiite generation as a precursor to sub-calcic garnet and diamond crystallization.

A suitable time differential between komatiite and diamond formation is also required to allow for dramatic radiogenic growth of ^{87}Sr in the enriched (high Rb/Sr and low Sm/Nd) precursor. In contrast, differential radiogenic growth of ^{143}Nd is far less pronounced because relative differences in low Sm/Nd ratio are small by comparison with those in Rb/Sr ratio. The magnitude of this time differential can be derived from Rb–Sr model age relationships, as illustrated in a $^{87}\text{Sr}/^{86}\text{Sr}$ evolution diagram (Fig. 4).

Hypothetical sub-calcic garnet host (diamond precursor) evolution curves are shown diverging from the bulk Earth curve towards the highest present day $^{87}\text{Sr}/^{86}\text{Sr}$ ratios measured in unencapsulated concentrate garnets. Intersections of the respective diamond inclusion garnet and precursor curves yield a common Rb–Sr model age for diamond formation of $\sim 3,200$ Myr, predicated on an ~ 300 Myr time differential between precursor formation (enrichment of komatiite residua) and diamond crystallization.

This treatment can be extended to argue against an origin of such diamonds soon after formation of the Earth¹³. Intersections of extrapolated diamond curves with precursor curves diverging from the bulk Earth at 4,550 Myr, give model ages of $< 4,000$ Myr (Fig. 4). Thus, even with precursor enrichment dating back to original accretion of the Earth, formation of such diamonds before 4,000 Myr is effectively precluded.

As previously indicated, concentrate garnets show a range of major element compositions (Fig. 1) and $^{87}\text{Sr}/^{86}\text{Sr}$ and $^{143}\text{Nd}/^{144}\text{Nd}$ ratios (Fig. 2) that extend towards those characteristic of peridotite xenolith garnets¹⁹. This suggests that original enrichment in sub-calcic garnet host assemblages has variably dispersed over time through open system behaviour. In principle and practice, more extreme concentrate garnets may remain to be analysed. Thus, precursor evolution curves may possibly have been even steeper than indicated, with a correspondingly smaller time differential between enrichment and diamond crystallization.

Nature of enriched precursor

The overall consistency of Sm–Nd and Rb–Sr model age relationships is taken as strong support for a 3,200–3,300 Myr origin of diamonds following enrichment of residues or high-pressure cumulates from 3,500 Myr komatiites. Given the restriction of sub-calcic concentrate garnets, diamonds and komatiites to the Archaean cratons in southern Africa, the most likely site of sub-calcic garnet and diamond formation is in upper mantle harzburgitic assemblages underlying the cratons¹¹. Sub-cratonic storage of such diamond host assemblages since that time implies that continental lithosphere was stabilized to depths within the diamond stability field during the Archaean. The existence of

a thick, cold lithosphere at that time is corroborated by equilibration temperatures below $1,150^\circ\text{C}$ (but with a total range of 900 – $1,300^\circ\text{C}$ and potential systematic uncertainties) derived from olivine and garnet inclusion pairs in diamonds^{3,20}. These temperatures have also previously been used to infer sub-solidus crystallization of diamond³. The striking absence of corresponding host peridotite xenoliths has, in turn, been attributed to the presence of disseminated magnesite which would cause disruption by decomposition during eruption^{11,30}. Now, an additional alkali-rich phase is apparently needed to account for sub-calcic garnet Rb–Sr systematics. However, neither magnesite nor an alkali host phase such as phlogopite is an obvious member of the peridotitic diamond inclusion suite. Therefore, the mechanism of enrichment is suggested to be introduction and entrapment of asthenosphere-derived alkali, LREE and CO_2 enriched interstitial melt, which remained liquid until the time of diamond crystallization². Furthermore, residual interstitial melt may have persisted through time, allowing for easy diffusive exchange with unencapsulated garnet, and assuring disaggregation during sampling by kimberlite².

Implications for kimberlite compositions

The inferred trace element and isotopic characteristics of sub-calcic garnet (and diamond) host assemblages make them attractive candidates for the enriched component apparent in micaceous kimberlites from southern Africa¹⁵, as well as kimberlites and lamproites from the Kimberley region of western Australia³¹. If kimberlites originate with distinctly depleted Sr and Nd isotopic characteristics analogous to those of ocean islands³² rather than those of a bulk Earth reservoir³³, the observed isotopic compositions of different kimberlite types may well reflect variable incorporation of such old enriched diamond host assemblages. This is in accord with the speculation of Nixon *et al.*³⁴ that kimberlites are hybrid magmas by virtue of incorporation of incompatible element enriched melt at the base of the continental lithosphere.

We are heavily indebted to De Beers Consolidated Mines for provision of all durable raw materials. J. B. Hawthorne, Consulting Geologist (Diamonds), and J. Hartley, A. van Niekerk and V. Anderson, the indefatigable sorters of the Diamond Sorting Office in Kimberley, are particularly acknowledged for making this project possible. We are similarly grateful to S. R. Hart for hosting this geochemical endeavour at MIT. K. Burrhus is credited with innovative diamond cracker design and, in partnership with S. R. Hart, paternity of NIMA-B. Professional assistance from J. Robey, D. Hall, G. Beye, A. Hurshman and S. Luria was appreciated. Financial assistance from the Harry Crossley Foundation (SHR), the University of Cape Town, the CSIR and NSF grants EAR78-03342 & EAR83-08809 (supporting SRH's isotope laboratory) is acknowledged.

Received 7 February; accepted 18 May 1984.

- Ringwood, A. E. *Earth planet. Sci. Lett.* **36**, 443–448 (1977).
- Harte, B., Gurney, J. J. & Harris, J. W. *Contr. Miner. Petrol.* **72**, 181–190 (1980).
- Boyd, F. R. & Finnerty, A. A. *J. geophys. Res.* **85**, 6911–6918 (1980).
- Gurney, J. J. & Harte, B. *Phil. Trans. R. Soc. A297*, 273–293 (1980).
- Meyer, H. O. A. & Tsai, H.-M. *Miner. Sci. Engng* **8**, 242–261 (1976).
- Sobolev, N. V. *Deep-Seated Inclusions in Kimberlites and the Problem of the Composition of the Upper Mantle* (Engl. transl. Brown, D. A.) (AGU, Washington DC, 1977).
- Harris, J. W. & Gurney, J. J. in *The Properties of Diamond* (ed. Field, J. E.) 555–591 (Academic, London, 1979).
- Sobolev, N. V., Galimov, E. M., Ivanovskaya, I. N. & Yefimova, E. S. *Dokl. Akad. Nauk SSSR* **249**, 1217–1220 (1979).
- Deines, P. *Geochim. cosmochim. Acta* **44**, 943–961 (1980).
- Gurney, J. J. & Switzer, G. S. *Contr. Miner. Petrol.* **39**, 103–116 (1973).
- Boyd, F. R. & Gurney, J. J. *Yb Carnegie Instn Washington* **81**, 261–267 (1982).
- Kramers, J. *Earth planet. Sci. Lett.* **42**, 58–70 (1979).
- Ozima, M., Zashu, S. & Nito, O. *Geochim. cosmochim. Acta* **47**, 2217–2224 (1983).
- Barrett, D. R. & Berg, G. W. *Phys. Chem. Earth* **9**, 619–636 (1975).
- Smith, C. *Nature* **304**, 51–54 (1983).
- Harris, J. W., Hawthorne, J. B. & Oosterveld, M. M. *Terra Cognita* **2**, 200 (1982).
- Gurney, J. J., Harris, J. W. & Rickard, R. S. in *Kimberlites, Diatremes, and Diamonds: their Geology, Petrology and Geochemistry* (eds Boyd, F. R. & Meyer, H. O. A.) 1–15 (AGU, Washington DC, 1979).
- Harris, J. W., Hawthorne, J. B. & Oosterveld, M. M. in *Kimberlites, Diatremes, and Diamonds: their Geology, Petrology and Geochemistry* (eds Boyd, F. R. & Meyer, H. O. A.) 27–41 (AGU, Washington DC, 1979).
- Richardson, S. H. thesis, Massachusetts Institute of Technology (1984).
- Shee, S. R., Gurney, J. J. & Robinson, D. N. *Contr. Miner. Petrol.* **81**, 79–87 (1982).
- Hart, S. R. & Brooks, C. *Contr. Miner. Petrol.* **61**, 109–128 (1977).
- Zindler, A., Hart, S. R., Frey, F. A. & Jakobsson, S. P. *Earth planet. Sci. Lett.* **45**, 249–262 (1979).
- Davis, G. L. *Yb Carnegie Instn Washington* **76**, 631–635 (1977).
- Shimizu, N. *Earth planet. Sci. Lett.* **25**, 26–32 (1975).
- Hamilton, P. J., Evenson, N. M., O'Nions, R. K., Smith, H. S. & Erlank, A. J. *Nature* **279**, 298–300 (1979).
- Hamilton, P. J., O'Nions, R. K., Bridgwater, D. & Nutman, A. *Earth planet. Sci. Lett.* **62**, 263–272 (1983).
- Viljoen, M. J., Viljoen, R. P. & Pearton, T. N. in *Komatiites* (eds Arndt, N. T. & Nisbet, E. G.) 53–79 (Allen and Unwin, London, 1982).
- Arndt, N. T. & Nisbet, E. G. (eds) *Komatiites* (Allen and Unwin, London, 1982).
- Green, D. H., Nicholls, I. A., Viljoen, M. J. & Viljoen, R. P. *Geology* **2**, 11–14 (1975).
- Wyllie, P. J., Huang, W.-L., Otto, J. & Byrnes, A. P. *Tectonophysics* (submitted).
- McCulloch, M. T., Jaques, A. L., Nelson, D. R. & Lewis, J. D. *Nature* **302**, 400–403 (1983).
- Kramers, J. D., Smith, C. B., Lock, N. P., Harmon, R. S. & Boyd, F. R. *Nature* **291**, 53–56 (1981).
- Basu, A. R. & Tatsumoto, M. *Science* **205**, 398–401 (1979).
- Nixon, P. H., Rogers, N. W., Gibson, I. L. & Grey, A. A. *Rev. Earth planet. Sci.* **9**, 285–309 (1981).
- Lawless, P. J. thesis, Univ. Cape Town (1978).
- Sobolev, N. V., Lavrentev, Y. G., Pokhilenko, N. P. & Usova, L. V. *Contr. Miner. Petrol.* **40**, 39–52 (1979).
- DePaolo, D. J. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **43**, 615–627 (1979).
- Allegre, C. J., Ben Othman, D., Polve, M. & Richard, P. *Phys. Earth planet. Inter.* **19**, 293–306 (1979).
- Jacobsen, S. B. & Wasserburg, G. J. *Earth planet. Sci. Lett.* **50**, 139–155 (1980).
- Wasserburg, G. J., Jacobsen, S. B., De Paolo, D. J., McCulloch, M. T. & Wen, T. *Geochim. cosmochim. Acta* **45**, 2311–2323 (1981).
- Minster, J.-F., Ricard, L.-P. & Allegre, C. J. *Earth planet. Sci. Lett.* **44**, 420–440 (1979).
- Allegre, C. J., Hart, S. R. & Minster, J.-F. *Earth planet. Sci. Lett.* **66**, 191–213 (1983).
- Jahn, B.-M. & Shih, C.-Y. *Geochim. cosmochim. Acta* **38**, 873–885 (1974).

From grasshopper to *Drosophila*: a common plan for neuronal development

John B. Thomas*, Michael J. Bastiani*, Michael Bate† & Corey S. Goodman*

* Department of Biological Sciences, Stanford University, Stanford, California 94305, USA

† Department of Zoology, Cambridge University, Downing Street, Cambridge CB2 3EJ, UK

Little is known about the mechanisms that generate neuronal specificity during development. Whereas the grasshopper embryo has been an ideal system for a cellular analysis of neuronal development, the Drosophila embryo has obvious attributes for a molecular genetic analysis. Here we show that the early Drosophila embryo is a miniature replica of the grasshopper embryo in terms of its identified neurones, their growth cones and their selective fasciculation choices, thus opening the way for a combined cellular and molecular genetic analysis of cell recognition during neuronal development in Drosophila.

BUILDING a functional nervous system involves the generation of enormous cellular diversity and at the same time the unfolding of remarkable neuronal specificity in terms of the precise patterns of interconnections among neurones. Elucidation of the mechanisms giving rise to neuronal specificity will require understanding at the cellular and molecular level of how individual neurones recognize and interact with one another during development.

The mechanisms underlying neuronal specificity are most easily studied in nervous systems containing relatively small numbers of neurones ($\sim 10^3$ neurones in an insect segmental ganglion), and in which the specificity is refined to such a degree that individual neurones can be uniquely identified according to their morphology and synaptic connectivity. A straightforward analysis of this question at the cellular level has been possible in the grasshopper embryo because of its relatively simple segmental ganglia, and most important because the embryonic neurones developing within these ganglia are large and highly accessible from their birth to maturation. Over the past decade, these attributes of the grasshopper embryo have permitted advances in our understanding of the cellular events of neuronal development (see, for example, refs 1–5). Moreover, these results have led to models for neuronal recognition which await testing and elucidation at the molecular level⁵.

Whereas the grasshopper embryo has been an ideal system for cellular studies, the *Drosophila* embryo has obvious attributes for a molecular genetic approach. The problem with studying cell recognition and neuronal specificity in the central nervous system (CNS) of the *Drosophila* embryo has always been the small size of its seemingly inaccessible embryonic neurones. Until recently, nothing was known of the detailed cellular events underlying neurogenesis in *Drosophila*, thus precluding a rational attempt at a molecular genetic approach to the problem.

Given our perspective of neurogenesis in the grasshopper embryo, we wondered if we could use the same methods (albeit in scaled-down versions) to elucidate the cellular events of neurogenesis in the *Drosophila* embryo. Here we show that the early fly embryo CNS is a miniature replica of the grasshopper embryo in terms of its identified neurones, their growth cones and their selective fasciculation choices when confronted with a common embryonic axonal scaffold. Furthermore, all arthropod nervous systems seem to be constructed using the same embryonic plan. Thus, we can now use the advanced genetics and molecular biology of *Drosophila* to extend our analysis of neuronal specificity to the molecular level.

Neurogenesis in the grasshopper

Like most animals, arthropods have bodies that consist of a series of repeated metameric units. The fundamental pattern of the arthropod CNS is a segmental chain of cephalic, thoracic and abdominal ganglia. The $\sim 2,000$ neurones in each thoracic (and ~ 500 in each abdominal) ganglion are generated during

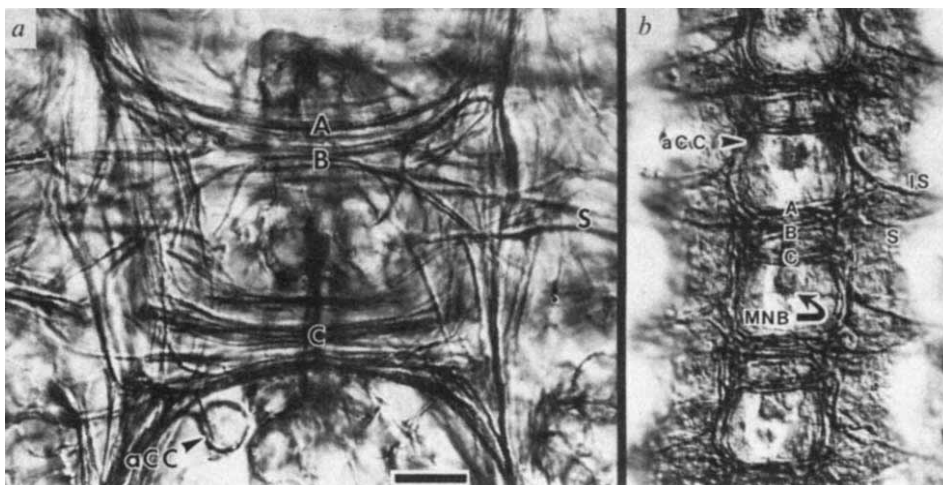
embryogenesis in the neuroepithelium from a repeated segmental pattern of neuronal precursor cells. With minor differences, each segment contains two bilaterally symmetrical plates of 30 neuroblasts (NBs)^{6,7}, an unpaired median neuroblast (MNB) and seven midline precursors (MPs)¹.

The set of neuronal precursor cells produces characteristic families of neurones; uniquely identified neurones are generated at specific points in the stem cell lineages. As each neurone differentiates, the stereotypic route taken by its growth cone can be charted^{8,9}. The early growth cones in the CNS interact with one another and use cues in the neuroepithelium to pioneer a stereotyped orthogonal scaffold of axon pathways, or fascicles (Fig. 1a). For example, the MP1 (one of the two progeny of midline precursor 1), and dMP2 and vMP2 (two sibling progeny of MP2) establish the first two longitudinal axon fascicles: the vMP2 fascicle and the MP1/dMP2 fascicle^{1,3,10}. The growth cones of the aCC and pCC neurones (siblings from the first division of NB 1–1) make cell-specific choices when confronted with these early axon fascicles: the pCC axon extends anteriorly and fasciculates on the MP1/dMP2 axons^{11,12} whereas the aCC axon waits and then extends posteriorly and laterally as it fasciculates on the later-developing U1 and U2 axons which pioneer the U fascicle (ref. 13 and S. DuLac and C.S.G., in preparation).

The growth cones of neurones differentiating later thus follow the axon bundles pioneered by earlier differentiating neurones. Each neurone's growth cone makes a series of cell-specific choices of which fascicle to follow, in this way generating the neurone's distinctive morphology. These stereotyped patterns of selective fasciculation are mediated by the high affinity that growth cones and their filopodia show for particular axonal surfaces. For example, in the 40% grasshopper embryo, the growth cone of the G neurone (from the second division of NB 7–4)^{14,15} is within filopodial grasp of ~ 25 different longitudinal axon fascicles (made up of ~ 100 different axons) and yet invariably chooses to fasciculate on the A/P fascicle (consisting of the A1, A2, P1 and P2 axons)¹⁶. When the A/P fascicle is ablated, the G growth cone does not show a high affinity for any other fascicle^{4,17}. Furthermore, extensive ultrastructural¹⁶ and experimental¹⁷ analysis demonstrates that G is able to distinguish the two P axons from the two A axons within the A/P fascicle; when only the P axons are ablated, the G growth cone behaves abnormally.

These examples of exquisite specificity have convinced us that many different molecules are differentially expressed on the surfaces of embryonic axon fascicles, or subsets of axons within them, and that they guide growing neurones to their appropriate targets by the selective adhesion of their filopodia to these labelled axonal pathways⁵. Monoclonal antibodies generated against the grasshopper neuroepithelium have revealed cell-surface antigens whose expression in the embryo correlates with the predictions of the cellular studies, namely, neurones

Fig. 1 Embryonic axonal scaffold of grasshopper (*a*) at 40% of development and *Drosophila* (*b*) at 12 h stained, respectively, with the I-5 monoclonal antibody²⁰ and an anti-tubulin monoclonal antibody (made by S. Loh, provided by T. Carr). A single segment which is the basic repeated unit of the scaffold is shown for grasshopper; three contiguous segments are shown for *Drosophila*, the commissures in one of which are labelled. The scaffold consists of bundles of axons arranged as two bilaterally symmetrical longitudinal connectives running the length of the embryo, three lateral commissures per segment joining the two sides of the ganglion (one posterior, C, and two fused anterior, A and B), and two peripheral pathways per segment, the intersegmental nerve (IS) at the segment boundary and the segmental nerve (S) at the segment midline (IS not included in *a*). In grasshopper there are ~25 axon fascicles in each of the connectives. Each of the three commissures contains ~20 different axon fascicles, many of which in the C commissure can be seen in the focal plane shown in *a*. This same pattern is present in *Drosophila*. Later-developing neurones grow along this framework, each choosing particular axon bundles on which selectively to fasciculate. In addition to the scaffold, several homologous cells can be identified in *Drosophila*: for example, the anterior corner cell (aCC) and the median neuroblast (MNB). Scale bar, 20 μ m for *a*, 15 μ m for *b*.



whose axons fasciculate together share common surface antigens¹⁸.

Neurogenesis in *Drosophila*

We now extend our study of neurogenesis to the molecular level by capitalizing on the advanced genetics and molecular biology of *Drosophila*. The difficulty with this approach has previously been the small size of the *Drosophila* nervous system and the lack of knowledge about the interactions and fates of individual embryonic neurones which develop within its tiny egg. There are some reasons for supposing that the grasshopper and the fruitfly might have radically different patterns of neural development. *Drosophila* is a holometabolous insect with two distinct phases of development: embryogenesis, which produces a larva, and metamorphosis, which transforms the larva during pupation into an adult fly. Embryogenesis in the more primitive (hemimetabolous) grasshopper, on the other hand, produces a larva which, in most respects, is a miniature adult requiring no radical metamorphosis. Adult structures such as limbs which are prefigured in the grasshopper embryo as limb buds are present in the *Drosophila* embryo and larva only as nests of undifferentiated cells, the imaginal disks.

However, although fruitflies and grasshoppers are separated by at least 300 Myr of evolution, we find that they, like the other arthropods we have studied, are constructed from a common plan for the embryogenesis of the arthropod CNS. The resemblance between *Drosophila* and the grasshopper is striking: it extends to individual identified neurones which appear homologous in both embryos (Figs 2–4), to the scaffold of axon fascicles upon which they grow (Fig. 1) and to the specific choices of selective fasciculation made by their growth cones (Fig. 5).

We have primarily focused on a comparison of neurogenesis in the fruitfly *Drosophila* and the grasshopper *Schistocerca*. However, in passing, we have also examined neurogenesis in two larger holometabolous insects (the moth *Manduca* and the fly *Calliphora*), in two other hemimetabolous insects (the bug *Rhodnius* and the cockroach *Periplaneta*) and in a crustacean (the crayfish *Procambarus*). With minor variations, the nervous systems of all of these insects (and one crustacean) are built from the same common embryonic plan, implying an early and subsequently conservative evolution of the developmental programme for the proto-insect (and perhaps proto-arthropod) CNS.

The *Drosophila* embryo is much smaller than the grasshopper embryo. We found, however, that we could dissect the living embryo in the same way as we do a grasshopper embryo. By

slitting the fly embryo down its dorsal surface and removing its gut, we reveal the dorsal surface of the neuroepithelium as a plate several hundred micrometres long, 50 μ m wide and 20 μ m thick. When viewed with Nomarski optics, the axon scaffold, neuronal cell bodies and in some cases individual axons (<0.5 μ m in diameter) can be visualized. In the *Drosophila* embryo, the cell bodies and axons are so small (cell bodies are typically 2–4 μ m in diameter as opposed to 10–15 μ m in the grasshopper) and so prone to damage from both the impalement and the UV illumination that the filopodia and even growth cones often retract and appear blebby. This problem can be overcome by lightly fixing the dissected embryos before dye injection (with 2% paraformaldehyde for 8 min). These lightly fixed embryos have intact plasma membranes and can be stored in buffer overnight. The fixed cells are taut and more easily penetrated. Although we do

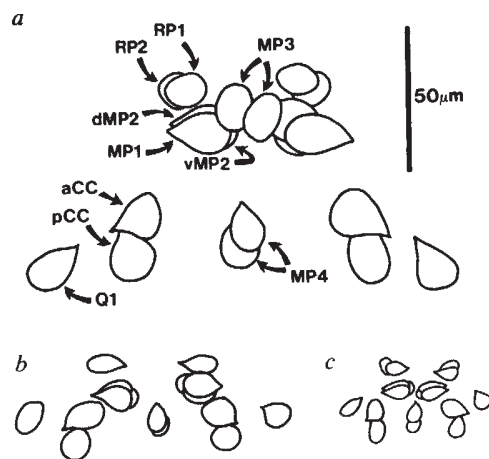


Fig. 2 Homologies of cell body identity and location of identified neurones in grasshopper (*Schistocerca*)(*a*), crayfish (*Procambarus*)(*b*) and fruitfly (*Drosophila*)(*c*). Embryos were dissected from their egg cases, filleted on a microscope slide and viewed with Zeiss Nomarski optics through a Leitz $\times 50$ salt water immersion objective. Camera lucida tracings were made of the identified cell body positions. Homologies were determined by each neurone's characteristic cell body location and its stereotyped axonal morphology as revealed by intracellular injection of Lucifer yellow. The pattern is identical in each species with the exception that the two MP3 progeny (the H cell and H-cell sib) are absent in the crayfish and *Drosophila*. RP2 lies directly underneath RP1 in the crayfish and thus is not shown in the drawing.

not record a resting potential, we can visually place the microelectrode tip in the cell body. Such fixed neurones can be filled by iontophoresis with Lucifer yellow or horseradish peroxidase (HRP) and show excellent integrity of the growth cones and filopodia; for electron microscopy, after dye injection the embryos are fixed for a second time and processed using routine procedures (Fig. 5b).

Neuronal precursor cells

All species examined use the same two kinds of neuronal precursor cells as in the grasshopper—neuroblasts (NBs) and midline precursors (MPs)—arranged in fundamentally the same pattern. In the three other species for which detailed precursor maps have been constructed (*Calliphora*, *Manduca* and *Rhodnius*), the fundamental arrangement of seven rows of NBs, a single median NB and medial MPs is conserved. Each species, however, has specific differences in the precise number of precursors, just as within a species there are specific differences between segments. No detailed map exists for *Drosophila* because some of the NBs are so small that they are often difficult to distinguish from neighbouring epidermal cells. However, our partial NB map for *Drosophila* indicates that they are arranged in the standard pattern of 7 rows and include more than 20 NBs per hemisegment.

Identified neurones

Certain identified neurones in the grasshopper embryo are particularly accessible because of some conspicuous feature such as the location of their cell bodies or the pathway of their growth cones. These diagnostic neurones (and their precursors) at 35% of development in the grasshopper embryo include the MP1, dMP2 and vMP2 (MP2), H cell and H-cell sib (MP3), aCC and pCC (NB 1–1), RP1 and RP2 (unknown NB lineage), Q1 (NB 7–4) and the two MP4s (MP4). Figure 2 compares this set of identified neurones in the grasshopper (Fig. 2a), crayfish (Fig. 2b) and *Drosophila* (Fig. 2c) embryos. The pattern is highly invariant, with the exception that the MP3 progeny (the H cell and H-cell sib) are absent in the crayfish and fly (interestingly, they are present in *Manduca*). Another diagnostic neurone at 40% of development in the grasshopper embryo is the G neurone (NB 7–4); this cell is also present in the crayfish, moth and fly.

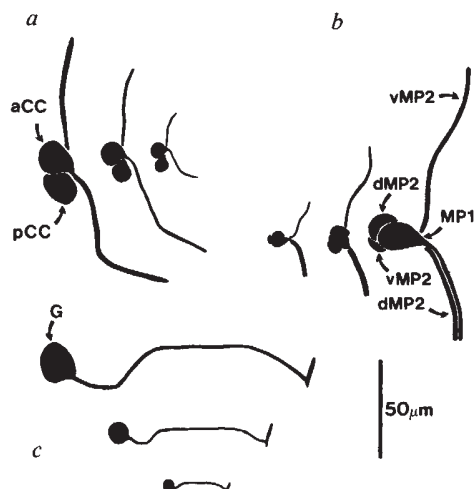


Fig. 3 Homologous pathway selection by identified neurones in (from largest to smallest in each set) grasshopper (*Schistocerca*), moth (*Manduca*) and fruitfly (*Drosophila*). The neurones compared for each species are the aCC and pCC (a), MP1, dMP2 and vMP2 (b) and the later-developing G neurone (c). Embryos were prepared and viewed as described in Fig. 2 legend; *Drosophila* embryos were fixed lightly in 2% paraformaldehyde for 8 min. Lucifer yellow was iontophoretically injected into the cell bodies; preparations were then viewed with fluorescence and camera lucida tracings made of the neurones.

These identified embryonic neurones also extend their growth cones along stereotyped pathways in all arthropod embryos examined. Figure 3 compares the morphologies of the aCC and pCC (Fig. 3a), MP1, dMP2 and vMP2 (Fig. 3b) and G neurones (Fig. 3c) in (from largest to smallest) the grasshopper, moth and fly. The *Drosophila* neurones seem to be simply miniature replicas of the grasshopper neurones.

Time course of neurogenesis

We have charted the development of highly accessible embryonic neurones in the fly embryo whose behaviours have been well documented in the grasshopper. Embryonic development in *Drosophila* takes 22 h (at 25 °C), compared with 20 days (at 33 °C)¹⁹ in the grasshopper. The cellular events of neurogenesis that occur over a 3-day period between 30% and 45% of development in the grasshopper, occur over a 3-h period during hours 10–13 in the fly. By 10 h the first neurones (MP1, dMP2, vMP2, pCC, aCC, RP1 and RP2) have begun to differentiate (the first growth cones appear at 9:45), extending axons which establish the first longitudinal, commissural and peripheral axonal pathways.

Figure 4 charts the development of these identified neurones in *Drosophila* at 10, 11 and 12 h of embryonic development. As in the grasshopper, the growth cones of MP1 and dMP2 extend posteriorly whereas the growth cone of the vMP2 extends anteriorly, pioneering a second longitudinal axon fascicle. By 10.5 h the MP1 and dMP2 growth cones have fasciculated with the pCC growth cone, extending anteriorly from the next posterior segment. By 11 h the MP1 growth cone has reached its homologue in the next posterior segment. The patterns of growth parallel those in the grasshopper, with minor modifications of temporal order. For example, in *Drosophila* the aCC extends its growth cone posteriorly and laterally to establish the intersegmental nerve simultaneous to the extension of the MP1, dMP2 and vMP2 growth cones to pioneer the first longitudinal fascicles. In grasshopper, the aCC growth cone does not extend posteriorly and laterally until some time after the MP1 and dMP2 growth cones have extended past it.

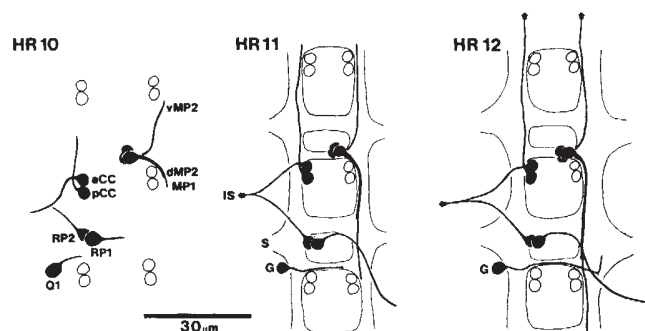
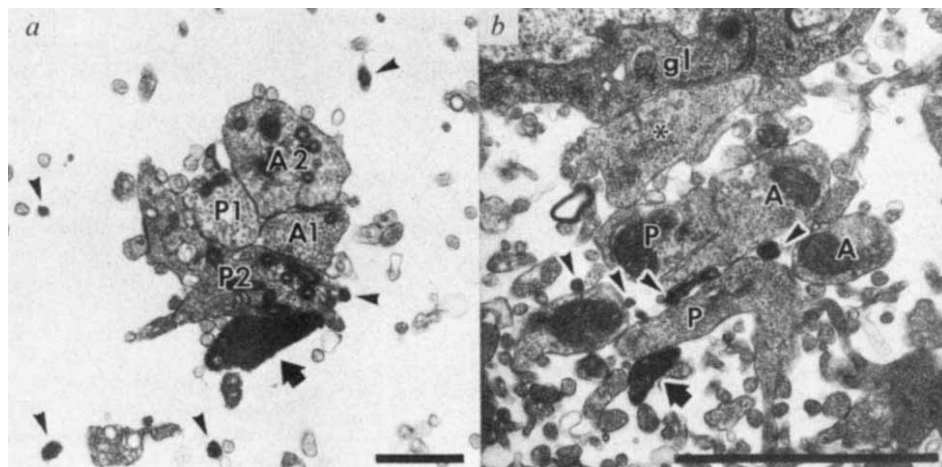


Fig. 4 Developmental time course of identified neurones in *Drosophila*. Wild-type O-R eggs were collected on yeast agar plates and allowed to develop at 25 °C. As there is some variation in developmental time between eggs within a single collection, each egg was staged on the basis of external embryo morphology after removal of the chorion²¹. Embryos were dissected from the vitelline membrane under saline, filleted on a microscope slide and lightly fixed for 8 min in 2% paraformaldehyde. Neurones were filled with Lucifer yellow via their cell bodies. Each developmental stage shown is a compilation of camera lucida drawings of Lucifer yellow fills from several embryos. Three contiguous segments are shown, marked by aCC and pCC cell bodies (open cell bodies). The first two axon fascicles in each of the longitudinal connectives are established by MP1, dMP2, pCC and vMP2 (see text). Cell Q1 (at 10 h) crosses the segment in what will become the posterior commissure (cc); the homologue of Q1 in grasshopper establishes the fascicle followed by its later-developing sibling, the G neurone. The G neurone in *Drosophila* crosses the segment between 11 and 12 h, grows past both the vMP2 fascicle and the MP1/dMP2 fascicle and selectively fasciculates with a more lateral bundle (the A/P fascicle) running within the connective. IS, intersegmental nerve; S, segmental nerve.

Fig. 5 Selective fasciculation in grasshopper and *Drosophila*. Transmission electron micrographs of the A/P fascicle and the tip of the G growth cone in a 40% grasshopper embryo (a) and a 12-h *Drosophila* (b) embryo from semi-serial TEM sections of HRP-filled G neurones. In both species the A/P fascicle contains the A1, A2, P1 and P2 axons and the G growth cone at this stage. In both species the tip of the G growth cone (large arrow) is in selective contact with the P axons rather than the A axons. Arrowheads show the filopodia of G. In *Drosophila* the axons are smaller in diameter and the fascicles closer together than in grasshopper. For example, the axon marked by an asterisk in b is probably homologous to the D fascicle, which in grasshopper is $\sim 10 \mu\text{m}$ dorsal to the A/P fascicle. Scale bar, $1 \mu\text{m}$.



Embryonic axonal scaffold

The early-differentiating neurones in the grasshopper embryo give rise to a stereotyped orthogonal pattern of axon fascicles which is repeated in each segment of the neuroepithelium. In the grasshopper embryo we can visualize this axonal scaffold using the I-5 monoclonal antibody²⁰ and HRP immunohistochemistry (Fig. 1a). The axonal scaffold at 40% of development in the grasshopper includes ~ 25 longitudinal axon fascicles on each side of the segment and ~ 20 axon fascicles in each of two commissures (one anterior which is subdivided into two commissures, A and B, and one posterior, commissure C) in each segment. Each segment also has two places where axon bundles leave the neuroepithelium and extend into the peripheral tissues—the intersegmental nerve (IS) at the segment boundary and the segmental nerve (S) at the segment midline.

This orthogonal scaffold of axon fascicles, constituting a ladder-like arrangement of longitudinal connectives, commissures and peripheral nerves along the dorsal surface of the neuroepithelium, is common to all arthropod embryos. In the *Drosophila* embryo, we can visualize the axonal scaffold using an anti-tubulin monoclonal antibody and HRP immunocytochemistry (Fig. 1b). The pattern of axonal fascicles represents a miniature version of the scaffold in the grasshopper embryo.

In the grasshopper embryo, the first two longitudinal axon bundles, the vMP2 fascicle and the MP1/dMP2 fascicle, are pioneered by the interactions of the MP1, dMP2, vMP2 and pCC neurones. The same is true in the fly.

Selective fasciculation

The detailed similarities between embryonic neurones in grasshopper and *Drosophila* suggest that each identified neurone differentiates in an equivalent environment within which the patterns of cell interactions, growth cone guidance and selective fasciculation are faithfully conserved. To substantiate this notion further, we have compared the development of the G neurone in *Drosophila* with that in the grasshopper, where much is already known about its selective fasciculation. As described above, in the 40% grasshopper embryo, the growth cone of the G neurone is within filopodial grasp of ~ 25 different longitudinal axon fascicles and yet invariably chooses to fasciculate on the A/P bundle consisting of the A1, A2, P1 and P2 axons. Furthermore, the tip of the G growth cone specifically fasciculates with the P and not the A axons (Fig. 5a). This stereotyped pattern of selective fasciculation by the G growth cone is mediated by the high affinity shown by its growth cone and filopodia for the surfaces of the P axons as compared with all other axons (~ 100) within filopodial grasp^{16,17}.

Drosophila has a neurone homologous to the grasshopper G

neurone (Figs 3, 4). After crossing the posterior commissure, its growth cone also turns anteriorly on a lateral bundle of axons in the contralateral neuropil. When this fascicle is reconstructed in the transmission electron microscope (TEM) after filling the G neurone with HRP, it is found to consist of two P axons and two A axons. Moreover, the tip of the fly G growth cone associates with the P and not the A axons, just as in the grasshopper (Fig. 5b). These and other results suggest that the patterns of selective fasciculation are largely identical in both insect species, and that the patterns of surface affinities and cell recognition have been highly conserved during insect evolution.

Discussion

The segmental ganglia in insects each contain $\sim 10^3$ neurones, many of which are unique identified neurones that have cell-specific morphologies and make cell-specific patterns of synaptic connections. This extraordinary degree of neuronal specificity is largely achieved during embryonic development by a sequential series of cell recognition events. The results of studies in the grasshopper embryo have led us to propose that cell recognition by neuronal growth cones is likely to involve the temporal and spatial expression of many different molecules⁵. Here we have shown that the same cellular events occur in the *Drosophila* embryo, although on a different temporal and spatial scale. In fact, this common theme for neurogenesis is faithfully observed in such diverse arthropod nervous systems as crayfish, grasshoppers and flies.

The most promising discovery is that by studying the large and accessible neurones of the grasshopper embryo for the past decade, we have in effect been studying the less accessible neurones of the *Drosophila* embryo. The fact that we can now work with identified neurones, their growth cones and their patterns of selective fasciculation in the *Drosophila* embryo means that genetic and molecular approaches can be applied to unravelling the mechanisms of cell recognition and neuronal specificity. The early cell recognition events involving selective fasciculation, so well characterized in the grasshopper embryo, occur between 10–13 h in the fly embryo. Fortunately, this is before the appearance of neurotransmitters, synapses and electrical excitability. Using recombinant DNA techniques we have begun to search for genes encoding the cell recognition molecules implicated by these cellular studies.

We thank David Hogness for the use of his *Drosophila* colony and Tim Carr for providing us with the monoclonal antibody against tubulin. This work was supported by a Helen Hay Whitney Postdoctoral Fellowship to J.B.T., a NICHD developmental biology postdoctoral training support grant to M.J.B., a grant from the Royal Society to M.B. and grants from the NIH, NSF and March-of-Dimes Birth Defects Foundation to C.S.G.

Received 19 March; accepted 18 June 1984.

1. Bate, C. M. & Grunewald, E. B. *J. Embryol. exp. Morph.* **61**, 317–330 (1981).
2. Goodman, C. S. & Bate, M. *Trends Neurosci.* **4**, 163–169 (1981).
3. Goodman, C. S., Raper, J. A., Ho, R. K. & Chang, S. *Symp. Soc. dev. Biol.* **40**, 275–316 (1982).
4. Raper, J. A., Bastiani, M. J. & Goodman, C. S. *Cold Spring Harb. Symp. quant. Biol.* **48**, 587–598 (1983).
5. Goodman, C. S., Bastiani, M. J., Raper, J. A. & Thomas, J. B. in *Molecular Bases of Neural Development* (ed. Cowan, W. M.) (Neuroscience Research Program Press, in the press).
6. Bate, C. M. *J. Embryol. exp. Morph.* **35**, 107–123 (1976).
7. Wheeler, W. M. *J. Morph.* **8**, 1–148 (1893).
8. Goodman, C. S. & Spitzer, N. C. *Nature* **280**, 208–214 (1979).
9. Goodman, C. S., O'Shea, M., McCaman, R. E. & Spitzer, N. C. *Science* **204**, 219–222 (1979).
10. Taghert, P. H., Bastiani, M. J., Ho, R. K. & Goodman, C. S. *Dev. Biol.* **94**, 391–399 (1982).

11. Bastiani, M. J. & Goodman, C. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1849–1853 (1984).
12. Bastiani, M. J. & Goodman, C. S. in *Cellular and Molecular Approaches to Neuronal Development* (ed. Black, I.) 63–84 (Plenum, New York, 1984).
13. Bastiani, M. J., DuLac, S. & Goodman, C. S. in *Model Neural Networks and Behavior* (ed. Selverston, A.) (Plenum, New York, in the press).
14. Raper, J. A., Bastiani, M. J. & Goodman, C. S. *J. Neurosci.* **3**, 31–41 (1983).
15. Raper, J. A., Bastiani, M. & Goodman, C. S. *J. Neurosci.* **3**, 20–30 (1983).
16. Bastiani, M. J., Raper, J. A. & Goodman, C. S. *J. Neurosci.* (in the press).
17. Raper, J. A., Bastiani, M. J. & Goodman, C. S. *J. Neurosci.* (in the press).
18. Kotrla, K. & Goodman, C. S. *Nature* (in the press).
19. Bentley, D., Keshishian, H., Shankland, M. & Toroian-Raymond, A. *J. Embryol. exp. Morph.* **54**, 47–74 (1979).
20. Chang, S., Ho, R. & Goodman, C. S. *Dev. Brain Res.* **9**, 297–304 (1983).
21. Bownes, M. *J. Embryol. exp. Morph.* **33**, 789–801 (1975).

DNA sequence and expression of the B95-8 Epstein–Barr virus genome

R. Baer*, A. T. Bankier, M. D. Biggin, P. L. Deininger*, P. J. Farrell†, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, C. Séguin*, P. S. Tuffnell† & B. G. Barrell

MRC Laboratory of Molecular Biology and † Ludwig Institute for Cancer Research, MRC Centre, Hills Road, Cambridge CB2 2QH, UK

The complete (172,282 base pairs) nucleotide sequence of the B95-8 strain of Epstein–Barr virus has been established using the dideoxynucleotide/M13 sequencing procedure. Many RNA polymerase II promoters have been mapped and the mRNAs from these promoters have been assigned to the latent or early/late productive virus cycles. Likely protein-coding regions have been identified and three of these have been shown to encode a ribonucleotide reductase, a DNA polymerase and two surface glycoproteins.

EPSTEIN–BARR virus (EBV) is a human herpesvirus¹ which is endemic in all human populations. Most people are infected with the virus in early childhood and then carry the virus for life. If the initial infection is delayed until adolescence, infectious mononucleosis frequently results. The virus is also linked with certain kinds of cancer. In the malarial belt of Africa, EBV is a contributory factor in the development of Burkitt's lymphoma and in South-East Asia the virus is linked to the high incidence of undifferentiated nasopharyngeal carcinomas (reviewed in refs 2, 3).

The lack of a simple permissive tissue system for EBV makes it difficult to obtain large amounts of virus and hampers genetic analysis. Until recently no genes had been located on the viral genome. Only primate B lymphocytes appear to have receptors for the virus^{4–6} but *in vivo* both B lymphocytes and certain (possibly epithelial) cells in the oropharynx become infected^{7,8}. There are many different EBV-infected lymphoid cell lines and these derive either from Burkitt's lymphoma explants or from B lymphocytes infected *in vitro* with EBV. When B lymphocytes are infected with EBV they are efficiently immortalized to perpetual growth. The B95-8 cell line was established by infecting marmoset B lymphocytes with EBV from a human with infectious mononucleosis⁹. Only a small proportion of the B95-8 cells support virus production spontaneously; the remainder are considered to be latently infected. The proportion of cells producing EBV can be substantially increased¹⁰ by adding various inducers such as the tumour promoter 12-*O*-tetradecanoylphorbol-13-acetate (TPA).

EBV has a double-stranded DNA genome about 172 kilobases (kb) long which is linear in the virus particle^{11,12} but exists as a circular episome inside the nucleus of the infected cell^{13,14}. The circularization is mediated by means of multiple direct sequence repeats about 0.5 kb long at the ends of the linear form^{15,16} which become joined in the circular form¹⁴. The genome is further divided into a short and a long unique region

by direct sequence repeats (up to 12) of ~3 kb^{17–19}. Unlike other herpesviruses, the short (*U_S*) and long (*U_L*) unique regions of EBV are maintained in a unique orientation relative to each other. In most EBV-containing cell lines (including B95-8) the majority of EBV DNA is episomal and it is generally thought that most gene expression in such lines is from this form.

The *EcoRI* and *BamHI* restriction fragments of the virus have been cloned and restriction maps for these enzymes obtained^{14,20,21}. By using these cloned fragments as probes in Northern blotting experiments, many viral mRNAs have been approximately localized on the viral genome^{22,23}. As a result of hybrid-selected translation experiments on mRNA from EBV-infected cells, many viral proteins have been assigned to regions of the viral genome^{24–26}. The genome is thought to be transcribed by the cellular RNA polymerases II and III.

We have now determined the complete DNA sequence of the B95-8 strain of EBV and analysed the sequence for likely coding regions and transcription promoters, splice junctions and poly(A) addition sites. This information is being used to analyse the transcription and gene expression of EBV. Here we show the overall arrangement of possible protein-coding regions and summarize our present knowledge of the transcription and translation of the virus.

DNA sequence analysis

M13 subclone libraries were constructed from suitable *EcoRI* or *BamHI* fragments of B95-8 EBV by the sonication method²⁷ using the M13mp8 and M13mp9 vectors²⁸. These M13 clones were sequenced randomly by the dideoxynucleotide procedure²⁹ and the data compiled in a DEC VAX computer using the programs of Staden³⁰. About 95% of the sequence of each region was obtained on both strands by the random procedure and the final single-stranded areas or ambiguities were resolved by more directed methods. The methods used have been reviewed by Bankier and Barrell³¹. The sequence of each nucleotide was determined on average 7.3 times and, apart from small parts of certain repetitive regions, determined on both strands. To join up the sequences of adjacent *EcoRI* or *BamHI* clones, random sequence analysis of a clone overlapping the junction was performed. Restriction maps deduced from the DNA sequence are shown in Fig. 1.

* Present addresses: Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA (R.B.); LSU Medical Center, 1901 Perdido St, New Orleans, Louisiana 70112, USA (P.L.D.); Cellular Regulation Section, Laboratory of Biochemistry, NCI, National Institutes of Health, Bethesda, Maryland 20205, USA (C.S.).

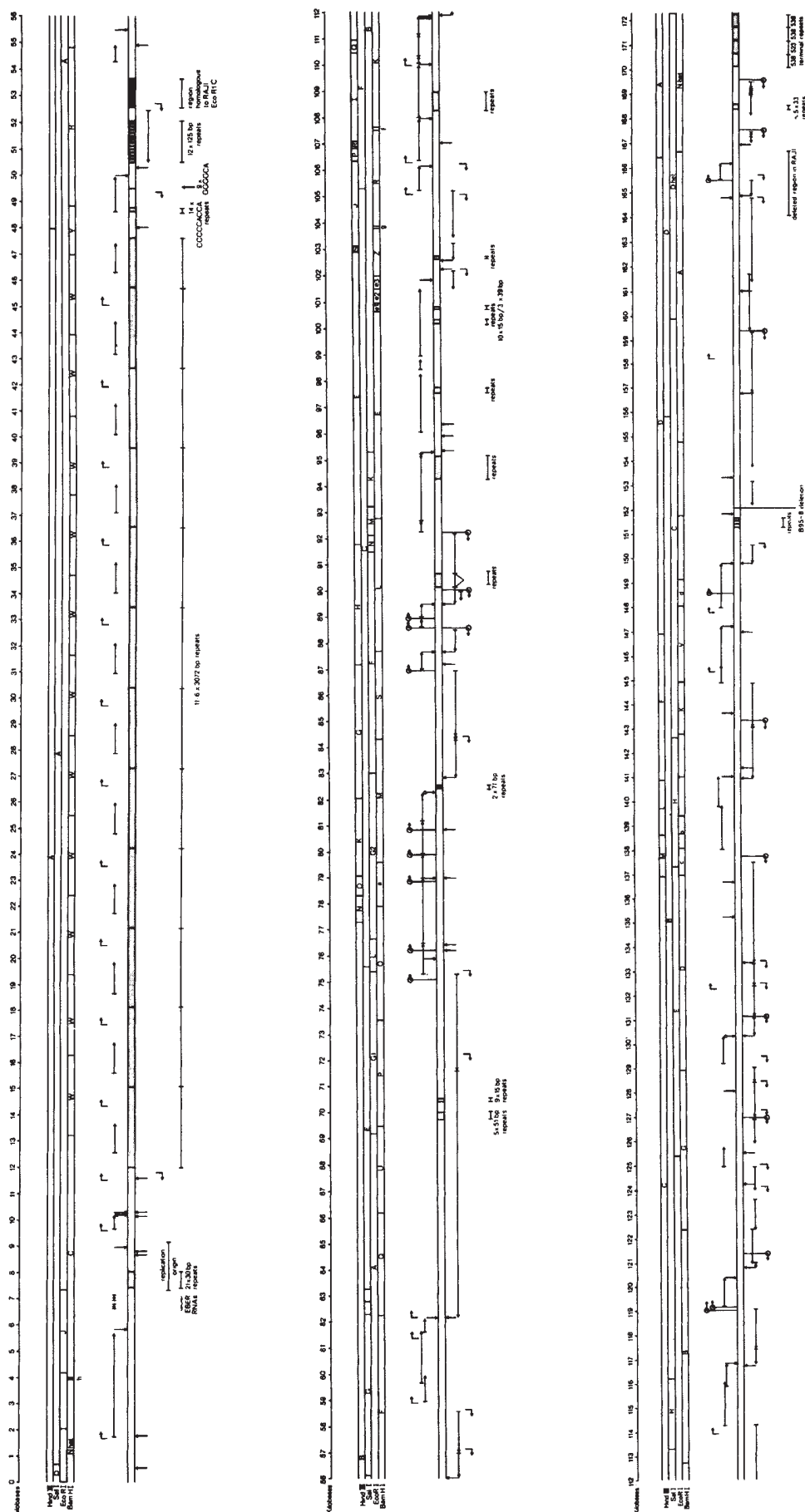


Fig. 1 Arrangement of restriction sites, major open reading frames, promoters and other features in the B95-8 EBV genome. Restriction maps deduced from the DNA sequence are shown for the enzymes *Hind*III, *Sal*I, *Eco*RI and *Bam*HI beneath the scale in kb. Position 0 on the scale is the *Eco*RI site between *Eco*RI Dhet and *Eco*RI I; this is 51 bases to the right of the true left end of the short unique region. Tandemly repeated regions of the genome are shaded. Reading frames are indicated by horizontal arrows and AATAAA (mRNA 3' ends and poly(A) sites) by vertical arrows. ? Marks the rarely used ATATAA 3' end signal. Promoters confirmed to function in B95-8 cells are shown by the symbol $\circ$ and promoters surmised from the DNA sequence are indicated by the symbol Δ . The reading frames shown in the diagram were selected on the grounds of length, codon usage, initiator methionine content and position relative to each other and to transcription signals. Confirmed promoters have been mapped by a combination of S₁ mapping and primer extension on RNA from B95-8 cells and in some cases by *in vitro* transcription. Possible promoters have been predicted from position and sequences homologous to the TATA box⁵³. The complete DNA sequence and feature table have been deposited with the EMBL Nucleotide Sequence Data Library, Postfach 10.2209, 6900 Heidelberg, FRG.

Table 1 Coordinates of the starts and ends of the reading frames indicated in Fig. 1 and the calculated molecular weights (MW) of the predicted polypeptides

Name	Position	MW	Promoter	Comments
BNRF1	1,736	5,689	142,843	
BCRF1	9,675	10,184	19,914	
BWRF1	12,541	13,689	39,866	12 copies
BYRF1	48,429	49,964	55,173	
BHRF1	54,376	54,948	21,893	
BHLF1	52,557	50,578	66,244	BH-L1 52,786 Early
BFLF2	56,935	55,982	35,361	
BFLF1	58,525	56,951	57,912	
BFRF1	58,891	59,898	37,632	
BFRF2	59,610	61,580	70,942	
BFRF3	61,456	62,034	19,966	
BPLF1	71,527	62,081	337,973	
BOLF1	75,239	71,523	132,750	
BORF1	75,238	76,329	39,191	BO-R1 75,052 Late
BORF2	76,407	78,884	93,030	BO-R2 76,198 Early
BarF1	78,900	79,805	34,358	Bar-R1 78,838 Early
BMRF1	79,899	81,110	43,373	BM-R1 79,871 Early
BMRF2	81,118	82,188	39,515	BM-R2 80,811 Late
BMLF1	84,122	82,746	51,347	
BSLF2	84,288	84,229	2,162	
BSLF1	86,881	84,260	98,040	
BSRF1	86,924	87,577	23,861	BS-R1 86,918 Late
BLLF2	88,474	87,641	30,952	BL-L3 88,481 Early
BLRF1	88,547	88,852	10,944	BL-R1 88,539 Late
BLRF2	88,925	89,410	17,687	BL-R2 88,896 Late
BLLF1a	92,153	89,433	94,431	BL-L1 92,158 Late
BLLF1b	92,153	89,433	75,171	gp350 gp220
BLLF3	90,013	89,569	16,652	BL-L2 90,020 Early
BLRF3	92,243	92,599	12,832	
BERF1	92,646	95,162	92,314	{ Homologous with BERF2b, BERF4
BERF2a	95,353	95,721	13,186	{ Homologous with BERF1, BERF4
BERF2b	95,725	98,244	92,769	{ Homologous with BERF1, BERF2b
BERF3	98,323	98,766	16,717	
BERF4	98,805	101,420	95,543	
BZLF2	102,116	101,448	25,257	
BZLF1	103,155	102,556	21,482	
BRF1	105,183	103,369	66,594	
BRF2	104,989	104,927	2,343	
BRRF1	105,182	106,111	35,319	
BRRF2	106,302	107,912	56,954	
BKRF1	107,950	109,872	56,427	EBNA
BKRF2	109,958	110,368	15,080	
BKRF3	110,275	111,117	31,606	
BKRF4	111,107	111,784	24,837	
BBLF4	114,259	111,833	89,853	
BBRF1	114,204	116,042	68,456	
BBRF2	115,843	116,781	34,916	
BBLF3	117,386	116,784	22,605	
BBLF2	119,080	117,416	60,364	
BBRF3	119,137	120,351	45,792	BB-R1 119,014 Late
BBLF1	120,974	120,750	8,470	BB-R3 119,133 Late
BGLF5	122,341	120,932	52,666	BB-L1 121,300 Late
BGLF4	123,692	122,328	51,291	
BGLF3	124,939	123,944	37,708	
BGRF1	124,938	125,912	36,462	
BGLF2	126,873	125,866	36,888	EE-L8 126,895 Late
BGLF1	128,374	126,854	54,462	
BDLF4	129,021	128,347	25,448	
BDRF1	129,188	130,348	42,626	
BDLF3	131,066	130,365	23,791	EE-L4 131,073 Late
BDLF2	132,389	131,130	46,168	
BDLF1	133,307	132,403	33,624	
BcLF1	137,466	133,324	153,916	EH-L1 137,676 Late
BcRF1	137,862	139,715	68,711	
BTRF1	139,642	140,916	46,711	
BXLF2	143,036	140,919	78,321	EC-L2 143,274 Late
BXLF1	144,861	143,041	67,193	
BXRF1	144,860	145,603	27,063	
BVRF1	145,416	147,125	62,461	
BVRF2	147,927	149,741	64,102	
BdRF1	148,707	149,741	36,127	EC-R1 148,651 Late
BILF2	150,525	149,782	27,076	
BILF1	153,099	152,164	34,519	
BALF5	156,746	153,702	113,419	
BALF4	159,322	156,752	95,640	EC-L1 159,337 Late
BALF3	161,678	159,312	85,536	
BALF2	164,770	161,387	123,122	
BALF1	165,517	164,858	25,149	
BARF1	165,504	166,166	24,471	ED-R1 165,498 Early
BNLF2b	167,303	167,001	11,449	
BNLF2a	167,486	167,307	6,540	ED-L2 167,495 Early
BNLF1c	168,966	168,163	28,851	
BNLF1b	169,129	169,043	3,212	
BNLF1a	169,474	169,208	9,942	ED-L1 169,517 Latent

BLLF1a and b are overlapping co-terminal reading frames, BLNF1b having a central portion spliced out. The locations of the approximate transcription start points of the confirmed promoters and the stage of expression of mRNAs from these promoters are given. Reading frames identified as coding for known proteins are indicated. Reading frames rich in the glycosylation site sequence (N-X-T/S) are also noted, together with a possible membrane protein, which has many hydrophobic amino acid residues.

The complete DNA sequence, together with a table showing the positions of the features shown in Fig. 1, has been deposited with the EMBL database; for reasons of space the sequence and feature table cannot be shown here. We have previously published the sequences of the *EcoRI* Dhet, *EcoRI* C, *BamHI* B, *BamHI* L, *BamHI* a and part of the *BamHI* M regions together with some analysis of their transcription³²⁻³⁸.

Most of the sequence we have established is from single *EcoRI* or *BamHI* clones. The possibility that cloning artefacts may have arisen in the construction and maintenance of the *EcoRI* and *BamHI* libraries cannot be excluded, but there is no reason to believe that such artefacts have occurred. The overlap between the *EcoRI* Dhet and *EcoRI* I fragments was obtained by sequencing the corresponding region of EBV strain, M-ABA, which has a single base change in the *EcoRI* recognition sequence.

There are some well documented areas of the EBV genome in which some strains have deletions relative to most other strains: one of these is in the *BamHI* WYH region and another is in the *EcoRI* C region. B95-8 does not have a deletion in the *BamHI* WYH region; the P3HR1 and Daudi strains, however, are missing bases 45,644 to 52,450 and 45,415 to 52,824 of our sequence respectively^{39,40}. There is a deletion of ~13.6 kb in B95-8 relative to most strains and we have determined that it lies between bases 152,012-152,013 by comparison with the sequence of the corresponding fragment of Raji DNA³³. Similarly, a deletion of 2,658 base pairs (bp) in the *EcoRI* Dhet fragment of Raji DNA removes bases 163,978 to 166,635 of the B95-8 sequence.

Repeat sequences

The genome is divided into the two unique regions (U_S and U_L) by the major internal repeat. We have avoided an earlier nomenclature in which the genome is subdivided into smaller unique regions by other repetitive sequences because there are many more repeat sequences in the virus than was previously thought and these do not appear to separate functional domains of the virus.

In the *EcoRI* Dhet clone that we sequenced (derived from the circular form of the virus) there were four copies of the terminal repeat, three of 538 bp and one of 523 bp³². In the 3.07-kb major internal repeat the sequence of only one of the *BamHI* W clones was determined and for the moment we have assumed that the repeats are identical. The number of copies of this repeat (11.6 copies in the B95-8 strain) has been taken from previous work⁴¹. Our sequence of *BamHI* W is identical to that determined by Jones and Griffin⁴² but differs in two places from the sequence of Cheung and Kieff⁴³.

Repetitive regions are scattered throughout the genome. Some repeats are found in likely coding regions and it is known in some cases that the repeats are actually encoded into protein: one of the most striking of these is the *BamHI* K Epstein-Barr nuclear antigen (EBNA) Gly-Ala repeat⁴⁴ in the BKRF1 reading frame (BKRF1 is explained below). In this repeat and several others there is no degeneration (third position variation) in the repeat sequences which are coding. Some mechanism or constraint apart from simply coding for protein may prevent these repeat sequences from drifting.

Interpretation of the sequence

The DNA sequence has been analysed⁴⁵ for transcription promoters, major open reading frames and possible polyadenylation/mRNA processing sites. The reading frames, promoters and AATAAA signals are shown in Fig. 1 together with other features of the sequence such as repetitive sequences. In our nomenclature, reading frames and promoters are preceded by the abbreviated name of the restriction fragment in which they start translation or transcription. A promoter starting transcription in *BamHI* K is preceded by BK-, one starting in *EcoRI* C is preceded by EC-. This is followed by L or R depending on whether the promoter or reading frame is leftward or rightward on the standard map. Promoters are then simply numbered;

reading frames have F and a number. Thus, BN-R1 would be a rightward promoter starting transcription in *Bam*HI N and BCLF2 would be a leftward reading frame beginning in *Bam*HI C. In many cases (see below) we have demonstrated promoters to be functional in B95-8 cells but some promoters are still only predicted from the DNA sequence.

On grounds of size alone it is likely that all or large parts of the reading frames shown are expressed as protein. Their positions with respect to each other and to promoter and polyadenylation/RNA processing sites strengthen this argument. Because of the high G+C content (59.94% in B95-8) we have been able to use codon usage analysis to further justify many of the reading frames³³. There are 84 unique major open reading frames shown in Fig. 1 and the coordinates and sizes of the reading frames and locations of known transcription starts and various other features are shown in Table 1. The sequence is numbered from the *Eco*RI site separating *Eco*RI Dhet from *Eco*RI I.

Generally the arrangement of coding regions in the unique regions is economical, particularly in the large unique region. Often there are few or no bases between reading frames and some promoters are found to lie in the coding regions of the adjacent genes. This economical arrangement in the unique regions contrasts with the large repetitive regions of the virus. The number of major internal repeat units is variable in different strains^{17,18,46,47} and tends to compensate for the deletions that certain strains have; this may indicate a packing constraint on the size of the virus.

Gene expression

The total number of proteins expressed from the EBV genome is unknown. Of the 84 major open reading frames, some may be spliced together, tending to reduce the total number of proteins expressed by the virus, while alternate splicing patterns would increase the number of proteins. Some proteins (discussed below) are thought to correspond to the major antigens of the virus, EBNA, MA (membrane antigen) and the capsid proteins (VCA). Other classically identified antigens such as the EA (early antigen) group and LYDMA (lymphocyte-detected membrane antigen), remain to be identified as polypeptide chains. A large number of EBV-specific polypeptides have been mapped to particular regions of the virus by hybrid-selected translation²⁴⁻²⁶, though in some regions of the virus no proteins have yet been detected by this method. For example, the two very large leftward reading frames BPLF1 and BOLF1 would encode polypeptides of molecular weight 338,000 and 133,000, respectively. However, no proteins have been reported as originating from this substantial region of the virus, though the fact that the reading frames are so large makes it fairly evident that they are expressed.

The EBV genome is thought to be transcribed by the cellular RNA polymerases II and III. RNA polymerase III transcribes the two Epstein-Barr early region (EBER) RNA genes in the small unique region^{48,49}. There is no evidence of any tRNA genes in EBV and our computer searches of the sequence have revealed none. The function of the EBER RNAs is not directly known but they will substitute for the adenovirus VA RNAs in an adenovirus infection⁵⁰. The EBER RNA genes are localized next to the maintenance origin of replication of EBV mapped by Yates *et al.*⁵¹.

We have searched for RNA polymerase II (pol II) promoters by a combination of identifying 5' ends of EBV mRNAs using S₁ mapping and primer extension experiments and by *in vitro* transcription of EBV DNA fragments in a HeLa whole-cell extract⁵². So far, 24 pol II promoters have been detected but the final total is expected to be more than twice this number. The location of the transcription starts of these promoters (which are all confirmed to function in B95-8 cells) are given in Table 1. All the promoters detected so far have sequences about 30 bases upstream of their transcription starts that are homologous with the TATAAAA sequence⁵³.

RNAs have been classified as expressed in the latent cycle or early or late productive cycles^{22,23}. We have compared the levels

of RNAs in B95-8 cells either not treated or treated with TPA. Although the control B95-8 cells produce virus, the levels of productive cycle RNAs are low in control cultures and increase very dramatically on TPA treatment so, in practice, it is easy to distinguish latent cycle RNAs from productive cycle ones in this system. Phosphonoacetic acid (PAA) is used as an inhibitor of viral DNA synthesis⁵⁴ to distinguish between early productive cycle RNAs and late productive cycle ones. PAA prevents the TPA induction of late RNAs but not of early RNAs.

It is unknown whether the regulated expression of EBV genes is controlled at the transcriptional level or through RNA processing or stability; there is a precedent for transcriptional regulation in herpes simplex virus (HSV)⁵⁵. We previously noted homologous DNA sequences upstream of some EBV promoters^{37,38} which might be used to regulate those genes at the transcriptional level. All the promoters which had those sequences give rise to late RNAs but only a few of the late promoters have the sequences, so it remains to be seen whether they are functionally important. The promoters are classified as latent, early or late in Table 1 though at present this classification refers to when the corresponding RNA is observed, rather than proven regulation of the promoter activity.

Genes active in latently infected cells

Significant levels of mRNA are expressed from three regions of the genome in latently infected cells in addition to the pol III EBER RNA transcripts^{23,56,57}: one of these regions is the *Bam*HI WYH region where mRNA is latently transcribed rightward possibly from the *in vitro* promoter⁵⁸ in *Bam*HI W (transcription start at position 45,104) or from the small unique region. A mature latent mRNA of 3.0 kb containing exons in *Bam* W, Y and H has been described⁵⁸. We have confirmed that there is a strong *in vitro* promoter starting transcription at position 45,104 but have been unable to show that it operates in B95-8 cells. Our preliminary Northern blotting experiments in the *Bam* W, Y, H region reveal two spliced rightward latent RNAs of 2.4 and 3.7 kb. The deletion in P3HR1 which may account for the nontransforming phenotype of this strain affects the region where these mRNAs map as well as the early RNA transcribed across the 125-bp repeats, encoding BHLF1³⁹.

A second latently transcribed region is *Bam*HI K. A 3.7-kb latent RNA hybridizes with *Bam*HI K which contains a simple repeat sequence 708 bases long, consisting only of the triplets GGA, GGG and GCA⁵⁹. This repeat sequence lies in the BKRF1 reading frame and codes for an amino acid sequence consisting of only Gly and Ala. This region has been conclusively shown to code for the nuclear antigen EBNA-1 (refs 44, 60). The molecular weight of the BKRF1 reading frame is 56,427 which is lower than the observed 68,000–85,000 of the EBNA-1 protein. The molecular weight of EBNA varies in a strain-specific manner according to the number of repeats in the BKRF1 frame⁶¹. The location of the promoter for latent transcription of the EBNA-1 gene is not yet known but may lie several kilobases upstream of the start of the BKRF1 frame⁵⁹. The observation⁶⁰ that a protein very similar to EBNA is produced in response to transfection of a *Bam*HI–*Hind*III fragment (positions 107,565–110,491) implies that BKRF1 accounts for most of the coding sequence. Another EBNA (EBNA-2) has been identified, though the gene for this has not yet been accurately localized on the genome map⁴⁴. A third nuclear antigen appears to map to the *Bam*HI M region⁶².

The most abundantly transcribed EBV mRNA in latently infected cells is a 2.8-kb RNA encoded by *Eco*RI Dhet^{23,56,57} which has been correlated with a latent active leftward promoter³⁸ at position 169,513 (ED-L1). In the latent virus cycle the RNA from this promoter is spliced and most of the mRNA is composed of an exon containing the BNLF1c reading frame and terminating at the poly(A) addition site at 166,950. This mRNA probably codes for a 42,000-molecular weight membrane protein³². The 5' end of this mRNA may be different in the productively infected cell.

Early productive cycle genes

Productive cycle RNAs are induced in B95-8 cells by treatment with TPA and induction of early RNAs is not inhibited by blocking viral DNA synthesis with PAA. The functions of three reading frames in B95-8 EBV have been identified by comparing the protein-coding sequences of all the reading frames in EBV with a library of known protein sequences. RNAs encoding these reading frames are expressed in the early productive cycle. We found that the reading frame BALF5 is similar to the HSV DNA polymerase (J. Quinn and D. McGeoch, personal communication) and the reading frames BORF2 and BaRF1 are similar (112 amino acids out of 301) to the HSV ribonucleotide reductase gene region³⁶. The EBV reading frame BXLF1 has a small region of homology (19 amino acids out of 35) with the HSV2 thymidine kinase gene⁶³. This sequence, however, may represent a nucleotide-binding site⁶⁴ rather than necessarily imply a thymidine kinase function for BXLF1.

Late productive cycle genes

The major component of the virus capsid is a 160,000 (160K)-molecular weight protein and a similar-sized protein has been observed by hybrid-selected translation using the *EcoRI* E fragment²². This protein may be encoded by the BcLF1 reading frame which would give a 154K protein.

The membrane antigen (MA) of EBV contains several proteins including gp350/300, gp250/200, p140 and gp85: three of these are glycoproteins (gp). Antibodies to MA and to the gp350/300 neutralize viral infectivity in tissue culture^{65,66}. The sequences of these proteins are of interest, therefore, in the development of synthetic vaccines against EBV infection. Hummel *et al.*⁶⁷ have mapped gp350/300 and gp250/200 to the *Bam*HI L region by hybrid-select translation and suggested that they may be expressed from overlapping reading frames because they have peptides in common. By Northern blotting and S₁ mapping³⁵, it has been shown that there are two co-terminal late RNAs containing the BLLF1 reading frame, the smaller having most or all of an internal repetitive region removed by splicing³⁵. It is proposed that gp350/300 is expressed from a 2.8-kb mRNA and gp250/200 is expressed from a 2.2-kb mRNA which is spliced, removing the repetitive region³⁵.

Although the location of the gene for the gp85 protein is unknown, two obvious candidates are the BALF4 and the BDLF3 reading frames as they are about the right size and contain many potential glycosylation sites. The non-glycosylated

membrane protein p140 is 143K in size and a similar-sized viral protein was mapped to the short unique region²²; this protein is probably encoded by BNRFL1.

Conclusions

Our structural approach to the biology of EBV, making use of the complete DNA sequence, has been particularly useful because of the lack of viral genetics and because of technical obstacles to working with the virus in tissue culture. The library of over 6,000 characterized M13 clones covering the genome obtained from the sequencing programme is proving most useful in analysing the gene expression of the virus. By using the M13 clones as probes for S₁ mapping and Northern blotting experiments, we are now constructing a detailed transcription map of the virus.

One of the most interesting features of EBV is its ability to immortalize B lymphocytes. Only a few regions of the genome are expressed in the latently infected lymphocyte: some of these may be involved in maintenance of the viral DNA and one at least is presumably involved in the immortalization. Identifying the immortalizing function may be useful both technically for making human monoclonal antibody lines that do not secrete EBV and with respect to the involvement of the virus in oncogenesis. EBV does not seem to be the proximal cause of Burkitt's lymphoma (reviewed in ref. 68) but is nevertheless a contributory risk factor. The fact that virtually every South-East Asian undifferentiated nasopharyngeal carcinoma carries EBV DNA implies a link between the virus and this disease. A region of the EBV genome which immortalizes epithelial cells has been identified⁶⁹ but immortalization of lymphocytes may require other EBV genes. At present it seems that an understanding of the role of the virus in these diseases will probably derive from a future investigation of the detailed molecular biology of EBV.

We thank B. E. Griffin and J. Arrand for the B95-8 and Raji *EcoRI* and *Bam*HI cloned restriction fragment libraries, and G. Bornkamm for the terminal M-ABA *EcoRI* clone. We also thank G. Bornkamm, B. E. Griffin, G. S. Hayward, S. D. Hayward, E. Kieff, G. Klein, D. J. McGeoch, G. Miller, M. Perricaudet, J. Pagano, J. Strominger, B. Sugden, H. Wolf and J. Yates who communicated their work to us before publication. During this study G.S.H. was supported by an Usher fellowship, P.L.D. by NATO, G.H. by the MRC, R.B. by the Damon Runyon-Walter Winchell Cancer Fund and C.S. by the MRC of Canada.

Received 28 February; accepted 1 May 1984.

- Epstein, M. A., Achong, B. G. & Barr, Y. M. *Lancet* **i**, 702-703 (1964).
- Rozman, B. (ed.) *The Herpesviruses, The Viruses* Vol. 1 (Plenum, New York, 1982).
- Epstein, M. A. & Achong, B. G. (eds) *The Epstein-Barr Virus* (Springer, Berlin, 1979).
- Henle, W., Diehl, V., Kohn, G., zur Hausen, H. & Henle, G. *Science* **157**, 1064-1065 (1967).
- Pope, J. H., Horne, M. K. & Scott, W. *Int. J. Cancer* **3**, 857-866 (1968).
- Jondal, M. & Klein, G. *J. exp. Med.* **138**, 1365-1378 (1973).
- Morgan, D. G., Niederman, J. C., Miller, G., Smith, H. W. & Dowlaty, J. M. *Lancet* **ii**, 1154-1157 (1979).
- Lemon, S. M., Hutt, L. M., Shaw, J. E., Li, J. L. H. & Pagano, J. S. *Nature* **268**, 268-270 (1977).
- Miller, G., Shope, T., Lisco, H., Stitt, D. & Lipman, M. *Proc. natn. Acad. Sci. U.S.A.* **69**, 383-387 (1972).
- zur Hausen, H., Hecker, E., O'Neill, F. J. & Freese, U. K. *Nature* **272**, 373-375 (1978).
- Pritchett, R. F., Hayward, S. D. & Kieff, E. *J. Virol.* **15**, 556-569 (1975).
- Hayward, S. D. & Kieff, E. *J. Virol.* **23**, 421-429 (1977).
- Lindahl, T. *et al. J. molec. Biol.* **102**, 511-530 (1976).
- Dambaugh, T. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 2999-3003 (1980).
- Given, D., Yee, D., Griem, K. & Kieff, E. *J. Virol.* **30**, 852-862 (1979).
- Kintner, C. R. & Sugden, B. *Cell* **17**, 661-671 (1979).
- Given, D. & Kieff, E. *J. Virol.* **31**, 315-324 (1979).
- Rymo, L. & Forsblom, S. *Nucleic Acids Res. S.* **5**, 1387-1402 (1978).
- Hayward, S. D., Nogee, L. & Hayward, G. S. *J. Virol.* **33**, 507-521 (1980).
- Skare, J. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3860-3864, 7510 (1980).
- Arrand, J. R. *et al. Nucleic Acids Res.* **9**, 2999-3014 (1981).
- Hummel, M. & Kieff, E. *J. Virol.* **43**, 262-272 (1982).
- Weigel, R. & Miller, G. *Virology* **125**, 287-298 (1983).
- Hummel, M. & Kieff, E. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5698-5702 (1982).
- Seibl, R. & Wolf, H. *Virology* (submitted).
- Cohen, L. K., Spech, S. H., Roberts, B. E. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Deininger, P. *Analyt. Biochem.* **129**, 216-223 (1983).
- Messing, J. & Vieira, J. *Gene* **19**, 269-276 (1982).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Staden, R. *Nucleic Acids Res.* **10**, 4731-4751 (1982); **12**, 499-503 (1984).
- Bankier, A. T. & Barrell, B. G. *Techniques in the Life Sciences* Vol. B5 (Elsevier, Ireland, 1983).
- Bankier, A. T. *et al. Molec. Biol. Med.* **1**, 425-445 (1983).
- Bankier, A. T., Deininger, P. L., Farrell, P. J. & Barrell, B. G. *Molec. Biol. Med.* **1**, 21-45 (1983).
- Seguin, C. S., Farrell, P. J. & Barrell, B. G. *Molec. Biol. Med.* **1**, 369-392 (1983).
- Biggin, M., Farrell, P. J. & Barrell, B. G. *EMBO J.* **3**, 1083-1090 (1984).
- Gibson, T., Stockwell, P., Ginsberg, M. & Barrell, B. G. *Nucleic Acids Res.* (in the press).
- Farrell, P. J., Deininger, P. L., Bankier, A. T. & Barrell, B. G. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1565-1569 (1983).
- Farrell, P. J., Bankier, A., Seguin, C., Deininger, P. & Barrell, B. G. *EMBO J.* **2**, 1331-1338 (1983).
- Jeang, K. & Hayward, S. D. *J. Virol.* **48**, 135-148 (1983).
- Jones, M., Foster, L., Sheedy, T. & Griffin, B. E. *EMBO J.* **3**, 813-821 (1984).
- Hayward, S. D., Lazarowitz, S. G. & Hayward, G. S. *J. Virol.* **43**, 201-212 (1982).
- Jones, M. D. & Griffin, B. E. *Nucleic Acids Res.* **11**, 3919-3937 (1983).
- Cheung, A. & Kieff, E. *J. Virol.* **44**, 286-294 (1982).
- Hennessy, K. & Kieff, E. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5665-5669 (1983).
- Staden, R., *Nucleic Acids Res.* **12**, 521-538 (1984).
- Raab-Traub, N., Pritchett, R. & Kieff, E. *J. Virol.* **27**, 388-398 (1978).
- Delius, H. & Bornkamm, G. W. *J. Virol.* **27**, 81-89 (1978).
- Arrand, J. R. & Rymo, L. *J. Virol.* **41**, 376-379 (1982).
- Rosa, M. D., Gottlieb, E., Lerner, M. R. & Steitz, J. A. *Molec. cell. Biol.* **1**, 785-796 (1981).
- Bhat, R. A. & Thimmappaya, B. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4789-4793 (1983).
- Yates, J., Warren, N., Reisman, D. & Sugden, B. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3855-3859 (1980).
- Corden, J. *et al. Science* **209**, 1406-1414 (1980).
- Summers, W. C. & Klein, G. *J. Virol.* **18**, 151-155 (1976).
- Herz, C. & Roizman, B. *Cell* **33**, 145-151 (1983).
- Kieff, E. *et al. J. infect. Dis.* **146**, 506-517 (1982).
- Rymo, L. *J. Virol.* **32**, 8-18 (1979).
- Van Santen, V., Cheung, A., Hummel, M. & Kieff, E. *J. Virol.* **46**, 424-433 (1983).
- Heller, M., van Santen, V. & Kieff, E. *J. Virol.* **44**, 311-320 (1982).
- Fischer, D. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 43-47 (1984).
- Hennessy, K., Heller, M., van Santen, V. & Kieff, E. *Science* **220**, 1396-1398 (1983).
- Grogan, E. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **80**, 7650-7653 (1983).
- Swain, M. A. & Galloway, D. A. *J. Virol.* **46**, 1045-1050 (1983).
- Walker, J. E., Sarate, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945-951 (1982).
- Hofmann, G. J., Lazarowitz, S. G. & Hayward, S. D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2979-2983 (1980).
- Thorley-Lawson, D. A. & Poody, C. A. *J. Virol.* **43**, 730-736 (1982).
- Hummel, M., Thorley-Lawson, D. A. & Kieff, E. *J. Virol.* **49**, 413-417 (1984).
- Klein, G. *Cell* **32**, 311-315 (1983).
- Griffin, B. E. & Karran, L. *Nature* **309**, 78-82 (1984).

Constraints on the neutron lifetime from triton β decay

J. Byrne

School of Mathematical and Physical Sciences,
University of Sussex, Brighton,
Sussex BN1 9QH, UK

The empirical data on triton β -decay, ${}^3\text{H} \rightarrow {}^3\text{He} + e^- + \bar{\nu}_e$, have been analysed with regard to their implications for neutron β decay, $n \rightarrow p + e^- + \bar{\nu}_e$. The latter process is a primary source of information on weak interactions in the nucleonic sector¹; it also has a central role in big-bang cosmology^{2,3} and in nuclear astrophysics in relation to the solar neutrino problem⁴. Unfortunately, the absolute rate of neutron β decay is rather uncertain⁵. The Particle Data Group has recommended the value $\tau_n = 925 \pm 11$ s for the neutron lifetime⁶, based on two direct measurements^{7,8} which are in close accord, but barely consistent with the value $\tau_n = 905 \pm 11$ s derived from directional correlation data on polarized and unpolarized neutron decay⁹⁻¹¹, and the comparative half lives of pure Fermi mirror transitions^{12,13}. A third direct measurement has given the value $\tau_n = 877 \pm 11$ s (ref. 14), which is totally at variance with all other evidence. We suggest here that the triton data exclude values of τ_n outside the range 911 ± 10 s, and deduce a weighted mean value $\tau_n = 914 \pm 6$ s.

Weak interactions are known¹⁵ to be mediated by vector bosons, in the same way that electromagnetic interactions are mediated by the exchange of photons. But whereas the space-time structure of the electromagnetic current is purely vector in character, the weak current has vector and axial components with corresponding form factors g_V and g_A analogous to the electromagnetic charge form factor. The weak current is further subdivided into leptonic and hadronic contributions, the latter having the transformation properties of an isovector. In the purely leptonic decay of the muon, g_V and g_A are identical, but this is not the case for neutron β decay, where the strong interactions are dominant. As an added complication, only the strangeness non-conserving part of the hadronic weak current contributes, resulting in an effective reduction of the weak coupling constant by a factor $\cos \theta_c$ (ref. 16). In the quark model of hadrons, the Cabibbo angle θ_c fixes the mixing of the d- and s-quark contributions in the hadronic weak current¹⁷.

In other respects, neutron β decay is very simple; there are no nuclear structure effects and the energy release $E_0 \approx 0.782$ MeV is so low that the energy dependence is manifested solely in the phase space factor f , which includes the effect on the Coulomb field of the residual proton (or indeed of a strong magnetic field, for example, inside a neutron star¹⁸). It is therefore convenient to express the decay rate in terms of the comparative half life $(ft)_n$, also referred to as the ft value, where $t = \tau \ln 2$ is the half life.

Triton β decay, like neutron decay, is a mixed transition between members of an isospin doublet. Because the mass excess of the second neutron in ${}^3\text{H}$ is closely matched by the Coulomb energy of the second proton in ${}^3\text{He}$, the two isospin substates are almost degenerate and the resultant low value of the energy release makes the transition ideal as a candidate for a neutron mass search^{19,20}. Thus, in view of the sustained interest in this question the main parameters of the decay have been determined with some precision.

From these new data we may calculate the ft value for triton decay and, subject to an important correction to the Gamow-Teller matrix element M_{GT} , a result may be derived for the number λ , which is defined as the ratio of the axial vector and polar vector form factors in neutron β decay. As the polar vector coupling constant is determined to an accuracy of $\sim 0.03\%$, either from the aforementioned Fermi decays or from the muon

lifetime²¹, and the Cabibbo angle is derived from weak decay rates in the baryon octet²², the neutron ft value and ultimately the neutron lifetime can be calculated.

Two values of the triton half life of comparable accuracy are currently available; the value $t_1 = 12.262 \pm 0.004$ yr (ref. 23) was determined from measurements of the growth rate of ${}^3\text{He}$ in a sample of molecular tritium, while the value $t_2 = 12.346 \pm 0.002$ yr (ref. 24) was obtained by calorimetry on samples of titanium titride and gaseous tritium. These results differ by $\sim 0.8\%$, a discrepancy which is well outside the claimed precision for either measurement but which is not of great importance in the present context.

We may calculate the phase space factor f using the parameterization of Wilkinson and Macefield²⁵ which takes account of nuclear recoil and radiative corrections up to $O(\alpha)$, and is accurate to better than 0.1% . Two recent measurements of the ${}^3\text{H}-{}^3\text{He}$ neutral atom mass difference, which agree on the value $18,573 \pm 7$ eV (refs 26, 27), give a result for the endpoint energy $E_0 = 18,532 \pm 7$ eV, allowing for the difference in binding energy of H and He^+ atoms²⁸. The resultant value for f is 2.8980×10^{-6} . Allowing for the $\sim 30\%$ of decays to excited atomic states²⁹, this number is reduced to 2.8913×10^{-6} ; on the other hand, inclusion of the 0.7% branch to neutral helium³⁰ and the 0.1% contribution of radiative corrections $O(\alpha^2)$ ³¹ tends to increase f . The final figure is $f = (2.914 \pm 0.006) \times 10^{-6}$.

We now exploit the relationship connecting ft values for allowed decays within $T = \frac{1}{2}$ and $T = 1$ isospin multiplets

$$(1 + |\lambda M_{GT}|^2) f t(\frac{1}{2}^+ \rightarrow \frac{1}{2}^+) = 2 f t(0^+ \rightarrow 0^+)$$

where the empirical average over Fermi mirror decays $\langle f t(0^+ \rightarrow 0^+) \rangle = 3,083.1 \pm 1.4$ s (ref. 5) provides, in effect, a measure of the polar vector coupling constant. From these data, we obtain the result $|\lambda M_{GT}|^2 = 4.455 \pm 0.022$, making due allowance for the uncertainty associated with the two competing values of the triton half life.

For the neutron $M_{GT} \equiv \sqrt{3}$, but $\lambda \neq 1$ (for the muon $\lambda = 1$) as the weak form factors are renormalized by the strong interactions. For the description of nuclear β decay in the impulse approximation, this renormalization is expressed by assigning effective weak charges to the nucleons, a procedure which is very successful for pure Fermi decays because the vector current is conserved. The axial current is only partially conserved (PCAC), and in Gamow-Teller transitions the axial coupling constant is quenched. This effect may be traced to core polarization and relativistic terms in the nuclear wave functions, and to meson exchange currents and in general is very complicated. However, in mirror nuclei with one hole or one particle outside an $N = Z$ core, the dominant nucleonic contribution to the core polarization vanishes exactly, and the remaining corrections now appear to be reasonably well understood³²⁻³⁵.

For the triton, Towner and Khanna³⁶ compute a net reduction $\delta = -2.6\%$ in M_{GT} , most of which (-1.6%) derives from the relativistic effect³⁷ because of almost complete cancellation of the core polarization and meson exchange contributions. Oset and Rho³⁸ concur and, by arguing that the quenching process may be understood in terms of the connection between the axial vector matrix element and the pion-nucleon scattering amplitude as expressed by the Goldberger-Treiman relation, find $\delta = -2.3 \pm 0.3\%$. To cover the full range of predictions, we shall assume that $\delta = -2.6 \pm 0.6\%$ giving the result $|M_{GT}|^2 = 2.846 \pm 0.035$. Combining the triton data, we find that $\lambda = 1.251 \pm 0.008$. The corresponding results for neutron β -decay are $(ft)_n = 1,083 \pm 12$ s and $\tau_n = 911 \pm 10$ s, where we have used the value $f_n = 1.71465 \pm 0.00015$ (ref. 5). This conclusion is consistent both with correlation data on neutron β -decay and with the direct lifetime measurements, excluding the one anomalous result¹⁴. We therefore ignore this result and average the combined triton and neutron data to give $(ft)_n = 1,086 \pm 7$ s with $\tau_n = 914 \pm 6$ s and $\lambda = 1.249 \pm 0.005$.

Finally we consider the theoretical significance of the parameter λ which, according to the Goldberger–Treiman relation, is given by the formula¹ $\lambda = f_\pi g_{\pi NN} / M_N$ where f_π is the charged pion decay constant, $g_{\pi NN}$ the pion nucleon coupling constant and M_N the mean nucleon mass. This relation is usually introduced in the context of PCAC theory but is currently viewed as a consequence of an exact chiral symmetry of the strong interactions whose associated conserved currents are identified with the strangeness conserving isovector and isoaxial weak currents. The non-existence of an isospin doublet, degenerate with the nucleons but of opposite parity, indicates that the isoaxial symmetry is spontaneously broken, and implies the

existence of an isotriplet of massless bosons with the quantum numbers of the pion³⁹. Because the pion is not, in fact, massless, the chiral symmetry is not exact, and the parameter $\Delta_\pi = 1 - \lambda M_N / f_\pi g_{\pi NN}$ provides a measure of the degree of symmetry breaking. Although it has proved difficult to compute an accurate value for this parameter^{40–42}, in a recent calculation based on current algebra and extended PCAC theory, Dominguez has obtained the fairly precise result $\Delta_\pi = 0.06 \pm 0.02$ (ref. 43). From the value of λ derived in the present work and using the data $f_\pi = 93.24 \pm 0.09$ MeV, $M_N = 938.926 \pm 0.002$ MeV and $g_{\pi NN} = 13.396 \pm 0.084$, we find the value $\Delta_\pi = 0.061 \pm 0.007$ which is in good agreement with the theoretical estimate.

Received 11 April; accepted 24 May 1984.

1. Byrne J. *Rep. Prog. Phys.* **45**, 115–184 (1982).
2. Taylor R. J. *Nature* **282**, 559–560 (1979).
3. Taylor R. J. *Q. Jl R. astr. Soc.* **24**, 1–9 (1983).
4. Bahcall J. H., Huebner W. F., Lubow S. H., Parker P. D. & Ulrich R. K. *Rev. mod. Phys.* **54**, 767–799 (1982).
5. Wilkinson D. H. *Nucl. Phys.* **A377**, 474–507 (1982).
6. Particle Data Group *Phys. Lett.* **B111**, 91–92 (1982).
7. Christensen C. J., Nielsen A., Bahnsen A., Brown W. K. & Rustad B. M. *Phys. Rev.* **D5**, 1628–1639 (1972).
8. Byrne J., Morse J., Smith K. F., Shaikh F., Green K. & Greene G. L. *Phys. Lett.* **B92**, 274–278 (1980).
9. Krohn V. E. & Ringo G. R. *Phys. Lett.* **B55**, 175–177 (1975).
10. Erozhinskii B. G., Frank A. I., Mostovoi Yu. A., Arzumanov S. S. & Voitzik L. R. *Soviet J. nucl. Phys.* **30**, 356–361 (1979).
11. Stratowa C., Dobrozemsky R. & Weinzierl P. *Phys. Rev.* **D18**, 3970–3979 (1978).
12. Wilkinson, D. H. & Alburgher, D. E. *Phys. Rev.* **C13**, 2517–2523 (1976).
13. Townner, I., Hardy, J. & Harvey, M. *Nucl. Phys.* **A284**, 269–281 (1977).
14. Bondarenko, L. N., Kurguzov, V. V., Prokofiev, Yu. A., Rogov, E. Y. & Spivak, P. R. *JETP Lett.* **28**, 303–307 (1978).
15. Arnison, G. *et al.* *Phys. Lett.* **B122**, 103–116 (1983); **B126**, 318–410 (1983).
16. Cabibbo, N. *Phys. Rev. Lett.* **10**, 531–533 (1963).
17. Glashow, S. L., Iliopoulos J. & Miani L. *Phys. Rev.* **D2**, 1285–1292 (1970).
18. O’Connell, F. R. & Matese J. J. *Nature* **222**, 649–650 (1969).
19. Bergkvist, K. E. *Nucl. Phys.* **B39**, 317–408 (1972).
20. Lyubimov, V. A., Novikov, E. G., Nozik, V. Z., Tret’yakov, E. G. & Kosik, V. S. *Phys. Lett.* **B94**, 266–268 (1980).
21. Duclose, J., Magnon, A. & Picard, J. *Phys. Lett.* **B47**, 491–493 (1973).
22. Roos, M. *Nucl. Phys.* **B77**, 420–428 (1974).
23. Jones, W. M. *Phys. Rev.* **100**, 124–125 (1955).
24. Jordan, K. C., Blanke, B. C. & Dudley, W. A. *J. inorg. nucl. Chem.* **29**, 2129–2131 (1967).
25. Wilkinson, D. H. & Macefield B. E. F. *Nucl. Phys.* **A232**, 58–92 (1974).
26. Simpson, J. J. *Phys. Rev.* **D23**, 649–662 (1981).
27. Smith, L. G., Koets, E. & Wapstra, A. H. *Phys. Lett.* **B102**, 114–115 (1981).
28. Bashkin S. and Stoner, J. R. *Atomic Energy Level and Grotrian Diagrams* Vol. 1 (North-Holland, Amsterdam, 1975).
29. Williams, R. D. & Koonin, S. E. *Phys. Rev.* **C27**, 1815–1817 (1983).
30. Bahcall, J. H. *Phys. Rev.* **124**, 495–499 (1961).
31. Wilkinson, D. H. *Nuclear Physics with Heavy Ions and Mesons* (eds Balian, R., Rho, M. & Ripka, G.) (North-Holland, Amsterdam, 1978).
32. Rho, M. *Nucl. Phys.* **A231**, 443–503 (1974).
33. Barroso, A. & Blin-Stoyle, R. J. *Nucl. Phys.* **A251**, 446–456 (1975).
34. Bahcall, J. H. *Nucl. Phys.* **A257**, 353–369 (1975).
35. Khanna, F. C., Townner, I. S. & Lee, H. C. *Nucl. Phys.* **A305**, 349–356 (1978).
36. Townner, I. S. & Khanna, F. C. *Phys. Rev. Lett.* **42**, 51–54 (1979).
37. Bell, J. S. & Blin-Stoyle, R. J. *Nucl. Phys.* **6**, 87–99 (1958).
38. Oset, E. & Rho, M. *Phys. Rev. Lett.* **42**, 47–50 (1979).
39. Pagels, H. *Phys. Rep.* **16**, 219–311 (1975).
40. Tegen, R., Schedl, M. & Weise, W. *Phys. Lett.* **B125**, 9–13 (1983).
41. Koniuk, R. & Tarrach R. Z. *Phys.* **C18**, 179–187 (1983).
42. Guichon, P. A. M., Miller G. A. & Thomas, A. W. *Phys. Lett.* **B124**, 109–112 (1983).
43. Dominguez, C. A. *Phys. Rev.* **D25**, 1937–1943 (1982).

NGC3256: an emerging elliptical galaxy

J. R. Graham, G. S. Wright, W. P. S. Meikle & R. D. Joseph

Astronomy Group, Physics Department, Imperial College of Science and Technology, Prince Consort Road, London SW7 2BZ, UK

M. F. Bode*

Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

NGC3256 is a spectacular peculiar galaxy, with a highly chaotic nuclear region consisting of several bright knots. Two diffuse tidal tails are visible evidence of violent past history. The galaxy has been classified on the basis of this tidal damage as the remnant of two colliding galaxies in the final throes of merging¹. The galaxy is highly luminous, with $M_B = -22.6$, as well as being a bright, narrow emission-line galaxy². At radio wavelengths, NGC3256 is one of the brightest in Wright’s survey³ of interacting galaxies. We have discovered bright and exceptionally extended 10- μ m emission that is evidence for an extremely luminous starburst extending over several kiloparsecs, which will leave the system severely gas-depleted. This depletion, in combination with the violent stellar relaxation which accompanies galaxy mergers, suggests that NGC3256 will become an elliptical galaxy.

NGC3256 was observed in February 1984 as part of our programme to investigate the IR properties of interacting galaxies⁴. Merging galaxies such as NGC3256 are an interesting subgroup of these, not only because the interactions are expected

to be particularly violent, but also because of the suggestion¹ that a significant fraction of elliptical galaxies could be the endpoint of mergers. The 1.5-m telescope at the Cerro Tololo Inter-American Observatory was used to make measurements at JHKL (1.25, 1.65, 2.2, 3.5 μ m) and 10 μ m in a 15 arc s aperture centred on the optical nucleus. 3 $\times$ 3 maps using the same beam, with pixel centres displaced one half beamwidth, were also made at these wavelengths. In addition, JHKL measurements were made in a 30 arc s aperture on the nucleus. At 10 μ m, the flux density measured on the nucleus was 1.7 ± 0.15 Jy. Emission was also detected in the pixels to the east and south-east at 1.2 ± 0.2 Jy and 1.3 ± 0.3 Jy respectively. The integrated flux of about 3 Jy makes this object one of the brightest extragalactic 10- μ m sources. At the distance of 53 Mpc inferred from its redshift⁵ ($H_0 = 50$ km s⁻¹ Mpc⁻¹) the nuclear 10- μ m luminosity of NGC3256 is $\approx 1.8 \times 10^{10} L_\odot$. This surpasses the 10- μ m luminosity of all recognized ‘starburst’ galaxies by an order of magnitude, and rivals that of the powerful Seyfert galaxies. Extrapolating this spectrum into the far IR suggests a total IR luminosity of order $3 \times 10^{11} L_\odot$ (ref. 6). The mass contained within the central 15 arc s (4 kpc) determined from the mean rotation curve⁵ is $6 \times 10^9 M_\odot$. Thus the mass/luminosity ratio in solar units is $\approx 0.02 (H_0/50)$. Clearly this ratio cannot be sustained by normal star formation over the lifetime of the galaxy.

The 10- μ m source must be extended on a scale of several arcseconds. This follows directly from consideration of the 10- μ m fluxes observed at three positions together with the measured beam profile. In fact, the 10- μ m emission is almost certainly more extended than these measurements suggest. A remarkably red $K - L$ colour ≈ 1.0 , indicating a large excess at L , extends over most of the central 30 arc s (typical galaxy colours are ≈ 0.3). Furthermore, one third of the total luminosity at L measured in the 30 arc s beam falls beyond the central 15 arc s. The most plausible interpretation of these results is that a mammoth burst of star formation has been triggered by the violent gravitational tides and collisional shocks produced by

* Present address: Department of Astronomy, The University, Manchester M13 9PL, UK.

the interaction. The steeply-rising IR spectrum looks very much like that typical of known starburst galaxies such as M82 (ref. 6). The very small mass/luminosity ratio, and the evidence for spatially extended 10- μ m emission are features characteristic of starbursts. In a starburst, the luminosity at 10 μ m and at L is due to massive early-type stars. The dust cocoon surrounding these stars at birth absorbs their light and reradiates it in the IR. The considerable variations we find in the *JHK* colours measured at different positions in our maps suggest that we have within the same aperture both the unreddened light from knots of less massive blue stars and emission from hot dust.

Apart from the intrinsic interest of an unusually luminous starburst, this discovery has one interesting further implication. There are both theoretical and observational arguments which support the suggestion that merging galaxies can evolve into ellipticals. Numerical simulations^{2,7} show that the violent relaxation following a collision produces a stellar velocity distribution similar to that in ellipticals. Observations of galaxies like NGC7257 and NGC5128 show this effect in the stellar velocity distribution, as well as an $r^{1/4}$ luminosity profile⁸. However, as White⁷ and others have pointed out, the ultimate fate of the gas which the two merging galaxies originally contain is an outstanding difficulty. If mergers generally result in powerful starbursts similar to that in NGC3256, this lacuna in the merger-elliptical evolution scenario is resolved. The massive early-type stars responsible for the high IR luminosity eventually become supernovae, and the mechanical energy in the supernovae ejecta is sufficient to drive a galactic wind which will leave the galaxy severely gas-depleted.

Galactic winds driven by supernovae have been discussed elsewhere⁹ but we can argue that, on simple energetic grounds, the supernovae resulting from the starburst in NGC3256 will be able to sweep the galaxy free of gas. The IR luminosity L_{IR} is powered by early-type stars of average luminosity l_* , so the total mechanical energy supplied by the supernovae produced when these stars explode is $(L_{\text{IR}}/l_*)E_{\text{SN}}$, where $E_{\text{SN}} \approx 10^{51}$ erg is the energy liberated by each supernova. The condition that this ejected gas, and any remaining interstellar gas, escape from the galaxy is

$$\frac{GM^2}{r} < \frac{L_{\text{IR}}}{l_*} E_{\text{SN}} \quad (1)$$

where we have made the conservative assumption that all the mass, M , inside a radius, r , is gas. In fact, a considerable fraction may be locked up in remnant objects and low-mass main sequence stars. If the duration of the starburst, τ_b , exceeds the lifetime of a typical early-type star, τ_* , then additional generations of supernovae will contribute to the kinetic energy, increasing it by a factor τ_b/τ_* . We may therefore re-express the condition for gas ejection in terms of the mass/luminosity ratio:

$$\frac{M}{L_{\text{IR}}} < 1 \times \left(\frac{\tau_b}{2 \times 10^7 \text{ yr}} \right) \left(\frac{m_*}{10 M_\odot} \right)^{-1} \frac{M_\odot}{L_\odot} \quad (2)$$

where m_* is the mass of a star characterizing the starburst. If we take a $10 M_\odot$ star to typify the early-type stars in the burst¹⁰, and we assume, conservatively, that there is only one generation of stars, (that which we are presently observing), then the condition for escape of all the gas is that $M/L_{\text{IR}} < 1$. As the observed M/L_{IR} for NGC3256 is 0.02, this condition is easily satisfied. Thus, there is more than enough mechanical energy to drive a galactic wind which will carry the gas out of the galaxy.

NGC3256 is, therefore, one of the most luminous examples of a starburst yet discovered. It also demonstrates that merger-induced star formation is sufficiently vigorous to produce a remnant as gas depleted as any elliptical galaxy. The discovery of the super-starburst in NGC3256 supports the idea that at least some of present-day elliptical galaxies could have been formed by mergers.

We thank the staff of the Cerro Tololo Inter-American Observatory, especially Jay Frogel, Brook Gregory and Oscar Rivera, for support. We also thank Norna Robertson for helpful

discussions. J.R.G. is a Department of Education for Northern Ireland supported research student. G.S.W. is an SERC supported research student. M.F.B. was supported by the Association of Universities for Research in Astronomy and by the US Department of Energy. J.R.G., W.P.S.M. and M.F.B. are visiting astronomers at Cerro Tololo Inter-American Observatory, which is operated by the Association of Universities for Research in Astronomy, Inc., under contract with the NSF.

Received 6 April; accepted 25 May 1984.

1. Toomre, A. in *Evolution of Galaxies and Stellar Populations* (eds Tinsley, B. M. & Larson, R. B.) 401–426 (Yale University Observatory, 1977).
2. de Vaucouleurs, G. & de Vaucouleurs, A. *Mem. R. astr. Soc.* **68**, 69–87 (1961).
3. Wright, A. E. *Mon. Not. R. astr. Soc.* **167**, 251–272 (1974).
4. Joseph, R. D., Meikle, W. P. S., Robertson, N. A. & Wright, G. S. *Mon. Not. R. astr. Soc.* (in the press).
5. Feast, M. W. & Robertson, B. S. C. *Mon. Not. R. astr. Soc.* **185**, 31–46 (1978).
6. Telesco, C. M. & Harper, D. A. *Astrophys. J.* **235**, 392–404 (1980).
7. White, S. D. M. *Mon. Not. R. astr. Soc.* **184**, 185–203 (1978).
8. Schweizer, F. in *Internal Kinematics and Dynamics of Galaxies*, IAU Symp. No. 100, 319–329 (1983).
9. Mathews, W. G. & Baker, J. C. *Astrophys. J.* **170**, 241–259 (1971).
10. Rieke, G. H., Lebofsky, M. J., Thompson, R. I., Low, F. J. & Tokunaga, A. T. *Astrophys. J.* **238**, 24–40 (1980).

γ Rays of 0.3–30 MeV from PSR0833–45

O. Tümay Tümer*, Bruce Dayton†, Jess Long*, Terrence O'Neill*, Allen Zych* & R. Stephen White*

* Institute of Geophysics and Planetary Physics, University of California, Riverside, California 92521, USA
† California State University, Los Angeles, California 90032, USA

Pulsed γ rays from the Vela pulsar PSR0833–45 are reported here for the first time at the medium energies, E , of 0.3–30 MeV. They were observed with the University of California, Riverside double scatter γ -ray telescope¹ flown on a balloon from Alice Springs, Australia, on 10 November 1981, 31 days after a large glitch in the pulsar period. The first and second pulses were detected from single scatters in the top scintillators, S1, at energies >0.3 MeV with statistical significances of 1.6σ and 1.6σ (together 2σ), and by double scatters from all detector cell pairs at energies of 1–30 MeV with significances of 4.8σ and 3.8σ (together 5.8σ). The phase separation of the two pulses is 0.43 ± 0.02 at the same absolute phases previously found by SAS 2² and COS B³ for $E > 35$ MeV. The energy distribution with six points from 0.3 to 30 MeV appears to bend away from the COS B power law⁴ at the lower energies and is well below the HEAO 1⁵ upper limits.

PSR0833–45 is the strongest γ -ray source in the sky^{2,3,6} for $E > 35$ MeV. Its energy distribution decreases as $E^{-1.89}$ from 0.05 to 5 GeV (ref. 4). It was discovered at the radio frequency of 408 MHz (ref. 7) and confirmed at other radio frequencies. Optical pulses were measured that are centred about the same phase as the two γ -ray pulses but only 0.25 in phase apart⁸. The single radio pulse leads the primary γ -ray pulse by 0.12 in phase. Upper limits, only, exist for soft and hard X rays⁵. The report⁹ of pulsed γ rays from 10 to 30 MeV with a narrow peak following the radio pulse has not been verified.

Our balloon was launched in the evening at 10.5 UT and reached an altitude of 4.5 g cm^{-2} at 15.5 UT when observations began, 4.5 h before the meridian passage of PSR0833–45. Observations continued for 9 h after zenith passage. In the double scatter mode the telescope measures the energy of the incident γ ray and its angle to a ring on the sky¹. Its fractional energy resolution varies from about 6 to 25% FWHM (half width at half maximum) for scatter angles in the first scintillator of 50° – 10° . The scatter angle resolution is $\sim 10^\circ$ FWHM independent of scatter angle. All cell pairs are used in the analysis. The

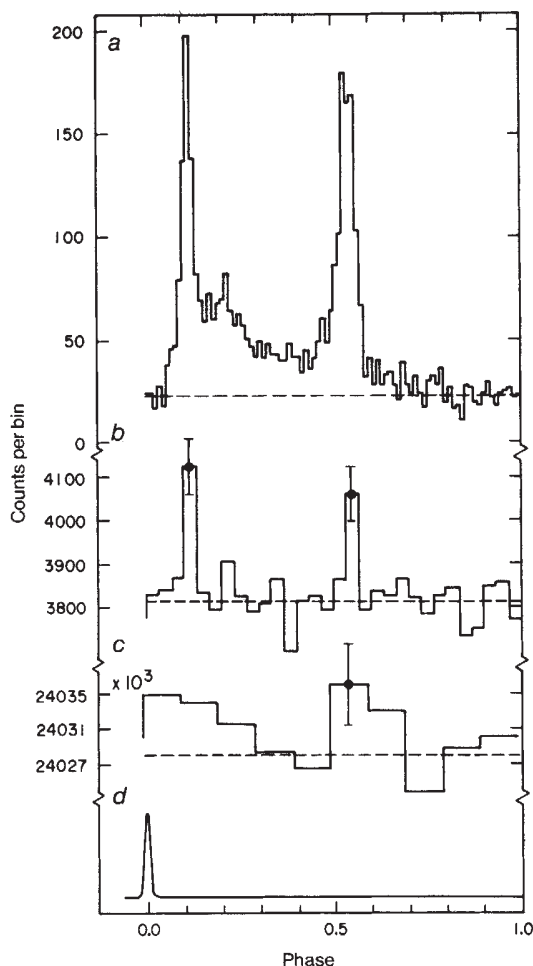


Fig. 1 Light curves for γ rays of: *a*, 50–6,000 MeV, COS B, ref. 12; *b*, double scatter, this work; *c*, 0.3–1.5 MeV, single scatter, this work timed absolutely relative to *d*, the 2,295 MHz radio light curve (R. N. Manchester, personal communication).

energy thresholds for the top (S1) and bottom (S2) scintillators were 0.32 and 0.40 MeV, respectively.

The Vela pulsar parameters and epoch provided by R. N. Manchester (personal communication) were determined from the seven central arrival times of the single radio pulse recorded on 3, 6, 9, 11, 16, 18 and 20 November 1981 at Honeysuckle Creek Radio Observatory, Australia. Manchester's values for the period P and $\dot{P}$ were verified with his topocentric arrival times and our program. The parameters are: epoch (Manchester's) 2444912.07142550356 JD; radio pulse phase 0.0; period 0.08925813256779 s; and period derivative $124.915107 \times 10^{-15} \text{ s s}^{-1}$. In the large glitch between 10 and 11 October 1981, Vela's period decreased by $\sim 102 \text{ ns}$ (ref. 10). Its recovery is accurately described by two exponential decays with time constants of 1.62 ± 0.18 and 233 ± 1 days. The P and $\dot{P}$ used for our observations 31 days after the glitch are in complete agreement with this observation¹⁰.

The direction of the source was determined from the overlapping of the rings on the sky. The total flux (pulsed plus constant) was determined for source bins on the sky as previously discussed¹¹. The contours for total flux showed a maximum in the direction of the Vela Pulsar¹¹.

The pulsed γ rays are analysed over the observation time of 15:46 UT 10 November to 0:50 UT 11 November 1981 with live time chosen for the Vela observation of 8 h 41 min 35 s (Vela zenith angle $< 59^\circ$). Selective angle cuts determined by the size of the cell combinations in S1 and S2 were applied along with the exclusion of events with cones $< 10^\circ$ above the horizon. γ -ray

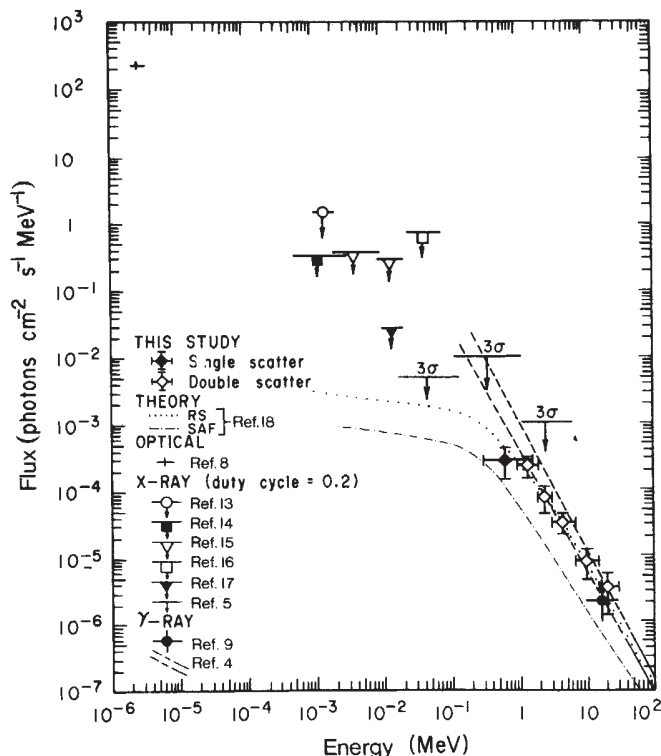


Fig. 2 The energy distribution of pulsed γ rays from the Vela pulsar PSR0833–45.

light curves obtained only from single scatters in S1 and from double scatters are shown in Fig. 1 along with the light curves for radio waves and for COS B γ rays¹² with $E > 50 \text{ MeV}$. For the single scatter mode, a count rate above a threshold of 0.3 MeV was recorded every 5.12 ms but γ -ray energies and directions were not measured for individual γ rays. The high efficiency area product of about $3 \times 10^3 \text{ cm}^2$ is responsible for the high count rates. Here the Vela zenith angle was kept $< 45^\circ$ to maximize signal-to-background noise. The uncertainty in phase of the single scatter light curve is ± 0.029 ($\pm 2.56 \text{ ms}$), half the accumulation bin size.

For double scatters, on the other hand, energies, angles and times to 0.05 ms were provided for each event. The dotted line shows the backgrounds calculated from the six channels not including the four channels in the pulses for the single scatter results and the 14 channels in the non-pulsed region between the second and first pulses in the double scatters. A χ^2 plot of a phase sweep $\pm 250 \text{ ns}$ from the Manchester pulse period and light curves at sampled periods confirm that the Manchester P was the best choice for our observations. The ratio of the counts in the inter-region between the first and second pulses to the counts in the first pulse, using the same background as for the pulses, is -0.1 ± 0.5 . This is to be compared with the value of about 0.8 ± 0.1 from COS B at energies of 50–300 MeV (ref. 12).

The absolute phases of the first and second pulses from the double scatters agree with those of SAS 2² and COS B¹² to our uncertainty of $< \pm 0.01$ in phase. Our phase separation between pulses 1 and 2 is 0.43 ± 0.02 , also in agreement. Our pulse widths of $0.045 \pm 0.015 \text{ FWHM}$ for each pulse, obtained from a 50-channel light curve, are in reasonable agreement with those of COS B for $E > 50 \text{ MeV}$.

A separate light curve was run for energies of 10–30 MeV. Our first and second pulses, seen with significances 2.5σ and 1.8σ (together 2.9σ), are located at the same phases as for Fig. 1. Our signal at the position of the single pulse previously reported at 10–30 MeV (ref. 9) is below background. We therefore find two pulses at 10–30 MeV with the same phases as at

our lower energies and at SAS 2 and COS B higher energies which are in disagreement with the previous single pulse.

The upper five energy points of our energy distribution come from the double scatters. Slightly different angle cuts were used in this analysis to match the efficiency calculations of the Monte Carlo program. The light curve of each of the energy intervals is plotted, each background is evaluated as above and the pulsed signal above background is taken from the same single channels (4 and 17) as for the first and second pulses in Fig. 1. The count rates are converted to fluxes and are plotted on Fig. 2 along with the COS B power law for $E > 50$ MeV and the upper limits at lower energies¹³⁻¹⁷. The dotted and dotted-dashed lines are the theoretical predictions for the shape of the energy distribution given by low-altitude acceleration polar cap models with large opening angles and pulsar radii of 10–30 km (ref. 18). The dotted curve is based on the Ruderman–Sutherland¹⁹, RS, model and the dotted-dashed curve on the Scharlemann *et al.*²⁰, SAF, model. The earlier point for 10–30 MeV with a single different phase pulse is also included. A least squares power law fit to the five points gives $(3.6 \pm 1.3) \times 10^{-4} E^{-(1.6 \pm 0.2)}$ photons $\text{cm}^{-2} \text{s}^{-1} \text{MeV}^{-1}$. This distribution was integrated between 0.3 and 1.5 MeV to estimate that 60% of the single scatters occurs between 0.3 and 1.5 MeV. A Monte Carlo calculated efficiency leads to the flux plotted in Fig. 2. Although the COS B energy distribution¹² best fit from 50 to 3,000 MeV follows the power law -1.89 ± 0.06 , the best fit from 50 to 300 MeV is -1.77 ± 0.15 . Our value of -1.6 ± 0.2 from 1 to 30 MeV continues the trend. All standard deviations and significances in this paper were calculated using equations (6) and (17), respectively, of ref. 21.

It appears that the points gradually bend away from the COS B curve at the lower energies and are well under the upper limits of HEAO 1. The power law, $2.2 \times 10^{-4} E^{-1.08}$ photons $\text{cm}^{-2} \text{s}^{-1} \text{MeV}^{-1}$, joining the 0.3–1.5 MeV point to the optical point⁸ at 3 eV does not violate the X-ray upper limits. The lower limit on the ratio of the energy in pulsed γ rays of 1–30 MeV to those of 1 eV–1 MeV is thus 10 while the ratio of the energy in γ rays of 1–30 MeV to those at higher energies of 30 MeV–3 GeV is about 0.5. Our energy distribution is in reasonable agreement with the predicted shapes of Fawley¹⁸ for the RS and SAF polar cap models. However, he also cautions that it will be difficult for the model to fit both the optical and γ -ray light curves. And the predicted energy distribution cuts off sharply about 1 GeV so it cannot explain a continuation of the COS B power law of $E^{-1.89}$ to higher energies.

We thank NSBF for their expert balloon launch, data retrieval and payload recovery, R. N. Manchester for providing the PSR0833–45 radio data and NASA for valuable support from grant 05-008-022.

Received 25 April; accepted 14 May 1984.

- Herzo, D. *et al. Nucl. Instrum. Meth.* **123**, 583–597 (1975).
- Thompson, D. J., Fichtel, C. E., Kniffen, D. A. & Ögelman, H. B. *Astrophys. J. Lett.* **214**, L17–L18 (1977).
- Bennett, K. *et al. Astr. Astrophys.* **61**, 279–284 (1977).
- Lichti, G. G. *et al. Adv. Space Explor.* **1**, 49–53 (1980).
- Knight, F. K., Matteson, J. L., Peterson, L. E. & Rothschild, R. E. *Astrophys. J.* **260**, 553–560 (1982).
- Wills, R. D. *et al. Proc. 17th Int. Cosmic Ray Conf.* **1**, 22–25 (1981).
- Large, M. I., Vaughan, A. E. & Mills, B. Y. *Nature* **220**, 340–341 (1968).
- Wallace, P. T. *et al. Nature* **266**, 692–694 (1977).
- Albats, P. *et al. Nature* **251**, 400–402 (1974).
- McCulloch, P. M., Hamilton, P. A., Royle, G. W. R. & Manchester, R. N. *Nature* **302**, 319–321 (1983).
- Tümer, O. T., Dayton, B., Long, J., O'Neill, T., Zych, A. & White, R. S. *Proc. 18th Int. Cosmic Ray Conf. XG4-1* (1983).
- Kanbach, G. *et al. Astr. Astrophys.* **90**, 163–169 (1980).
- Zimmerman, H. U. *Astr. Astrophys.* **88**, 309–310 (1980).
- Harnden, F. R. Jr *et al. Bull. Am. astr. Soc.* **11**, 424 (1979).
- Pravdo, S. H. *et al. Astrophys. J. Lett.* **208**, L67–L69 (1976).
- Ricker, G. R. *et al. Astrophys. J. Lett.* **186**, L111–L113 (1973).
- Yentis, D. J. *et al. in X-Ray Astronomy* (eds Baity, W. A. & Peterson, L. E.) 345–348 (Pergamon, Oxford, 1979).
- Fawley, W. M. thesis, Univ. California, Berkeley (1978).
- Ruderman, M. & Sutherland, P. *Astrophys. J.* **196**, 51–72 (1975).
- Scharlemann, E. T., Arons, J. & Fawley, W. M. *Astrophys. J.* **222**, 297–316 (1978).
- Li, Ti-pei & Ma, Yu-gian *Astrophys. J.* **272**, 317–324 (1983).

Evidence for haematite particles at 60 km altitude

Craig F. Bohren & John J. Olivero

Department of Meteorology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

Ackerman *et al.*¹ have reported experimental evidence for the existence of a dust layer at an altitude of 60 km (see ref. 2). Using their extinction measurements at 0.44 and 0.65 μm , they concluded that these dust particles are “made out of brownish matter”. They did not, however, speculate on what this matter might be. Nor did they estimate the size of the particles, other than that they are much smaller than the wavelengths of visible light. Our analysis, using their data, suggests that the particles are composed of haematite ($\alpha\text{-Fe}_2\text{O}_3$) and have radii around 200 Å. Experimental evidence either to support or to refute this condition could be provided by measurements in the UV, especially near 0.34 and 0.40 μm .

Extinction (scattering plus absorption) by sufficiently small absorbing particles is dominated by absorption. The absorption cross-section (normalized by the geometrical cross-section) of a small homogeneous sphere of radius a is approximately³

$$Q_{\text{abs}} = 4x \cdot \text{Im} \left\{ \frac{m^2 - 1}{m^2 + 2} \right\}$$

provided that $x \ll 1$ and $|m|x \ll 1$, where the size parameter x is $2\pi a/\lambda$, $m = n + ik$ is the complex refractive index of the particle (in air) at the wavelength λ and $\text{Im}\{\dots\}$ is the imaginary part of the quantity in parentheses.

We calculated Q_{abs} at 0.44 and 0.65 μm , using unpublished measured values of n and k for haematite ($\alpha\text{-Fe}_2\text{O}_3$). (The wavelength dependence of k for haematite and several solids and liquids that are found as atmospheric particles is shown in Fig. 14.1 of ref. 3.) Haematite is anisotropic, but not to the extent that this needs to be taken into account in an estimate of extinction by small haematite particles. Accordingly, we have used n and k appropriate to an electric field perpendicular to the optic axis.

At $\lambda = 0.65 \mu\text{m}$ (red) the complex refractive index of haematite is $3.08 + i0.13$; at 0.44 μm (blue) it is $3.2 + i1.35$. From the preceding expression for Q_{abs} it therefore follows that extinction of blue light exceeds that of red by a factor of about 11.6. This is consistent with the assertion by Ackerman *et al.* that “the optical efficiency of the layer increases more than 10 times when the wavelength of the interacting light changes from 0.65 to 0.44 μm ”.

Exact calculations of the ratio of absorptions at 0.44 and 0.65 μm are shown in Fig. 1. For radii ≈ 200 Å this absorption ratio is independent of size. As the radius is increased, however, the absorption ratio rises to a maximum at ~ 450 Å before a sharp decrease. This results from the greater contribution of the magnetic dipole term (and higher-order terms) in the series for the cross-section at 0.44 than at 0.65 μm , because k is much larger at the shorter wavelength; for very small particles, only the electric dipole term contributes to the cross-section. For sufficiently large particles we expect the ratio of absorption cross-sections to approach an asymptotic value < 1 , as indeed it does. This is easy to explain using geometrical optics arguments: all the incident light that is not externally reflected by a large absorbing particle enters it and is absorbed; according to the Fresnel formulae, however, reflection increases with increasing k .

Particles smaller than ~ 200 Å are consistent with the extinction observations of Ackerman *et al.*, but so, it could be argued from Fig. 1, is a distribution of particle sizes, some smaller than ~ 200 Å, some several times larger. Indeed, it could even be argued that the extinction observations equally support the assertions that the particles are smaller than 200 Å or that they have radii narrowly distributed around 700 Å.

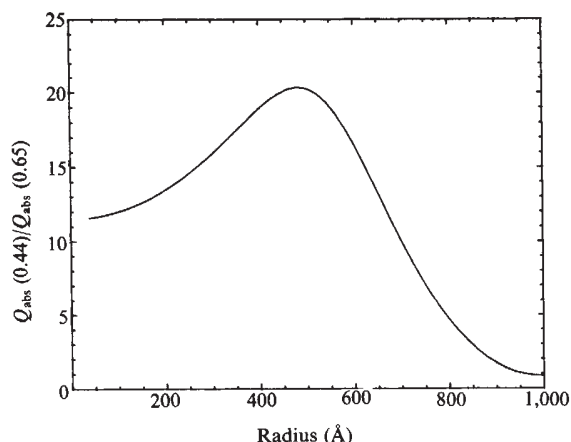


Fig. 1 Ratio of absorption cross-sections at 0.44 and 0.65 μm for haematite spheres of varying radius.

The small values of both the single-scattering albedo and the asymmetry parameter reported by Ackerman *et al.* support the hypothesis that the particles are smaller than 200 $\text{\AA}$. For example, a small ($\ll 1$) asymmetry parameter is a particularly good indicator of very small particles. As the size parameter is increased from values $\ll 1$ to values $\gg 1$, the asymmetry parameter rapidly increases nearly monotonically from zero to about 0.7–0.9. We shall therefore continue our analysis under the assumption that simple Rayleigh theory is adequate.

Ackerman *et al.* reported the single-scattering albedo ω_0 (ratio of scattering to extinction) at 0.44 μm . From the definition of ω_0 it follows that

$$\frac{\omega_0}{1 - \omega_0} = \frac{\langle a^2 Q_{\text{sca}} \rangle}{\langle a^2 Q_{\text{abs}} \rangle}$$

where Q_{sca} is the normalized scattering cross-section and the brackets indicate averages over the size distribution. For sufficiently small particles Q_{sca} is given by

$$Q_{\text{sca}} = \frac{8}{3} x^4 \left| \frac{m^2 - 1}{m^2 + 2} \right|^2$$

Size distributions of particles condensed from meteor ablation calculated by Hunten *et al.*⁴ are narrow. If we assume that such narrow distributions are applicable to the haematite particles under consideration here, the various moments $\langle a^n \rangle$ of the size distribution may be approximated by $K_n \langle a \rangle^n$, where K_n is not appreciably different from one. We therefore omit brackets from quantities, it being understood that x , for example, represents an average over the size distribution.

The value $\omega_0 = 0.1$ at 0.44 μm measured by Ackerman *et al.* implies that $x = 0.31$ (this is somewhat larger than the strict validity of the Rayleigh approximation dictates, but not by so much that it gives large errors). In turn, this size parameter implies a particle radius of $\sim 220 \text{\AA}$.

There is one other quantity, the asymmetry parameter g (mean cosine of the scattering angle), inferred by Ackerman *et al.* from their measurements. It may be shown from results in ref. 3 that to terms of order x^2

$$g = \frac{x^2}{15} \text{Re} \left\{ \frac{(m^2 + 2)(m^2 + 3)}{2m^2 + 3} \right\}$$

If we use the value $x = 0.31$ at $\lambda = 0.44 \mu\text{m}$ inferred from the single-scattering albedo measurements, this gives an asymmetry parameter of 0.050, which agrees well with the value 0.060 reported by Ackerman *et al.*

Results obtained using Mie theory and the calculated size distributions of Hunten *et al.*⁴, which we shall report in more detail elsewhere, are consistent with our conclusions about the size and composition of the particles observed by Ackerman *et al.*

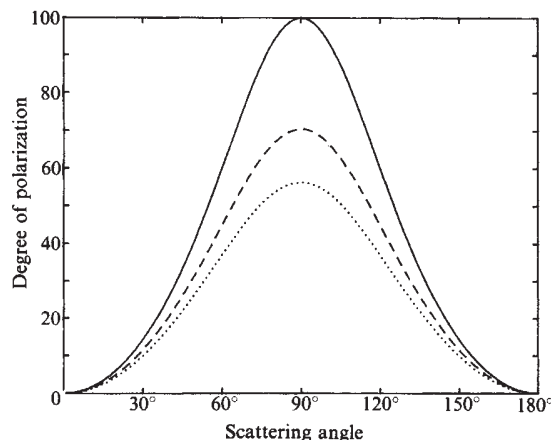


Fig. 2 Degree of linear polarization acquired by initially unpolarized light upon scattering by small (Rayleigh limit) spheres (—), randomly oriented disks (---), and randomly oriented needles (···). The refractive index is $3.08 + i0.13$.

We have assumed spherical particles and have analysed their optical properties using the Rayleigh approximation. If they are nonspherical, but still sufficiently small to be treated adequately by this approximation, the ratio of extinctions at the two wavelengths will be even greater than it is for spherical particles. Clues about the shape of small haematite particles may be obtained from polarization measurements. The degree of linear polarization of light scattered by such particles depends strongly on their shape because the refractive index is so high. This is shown in Fig. 2, where we have plotted the degree of polarization as a function of scattering angle for three ensembles of small haematite particles: spheres, randomly oriented disks and randomly oriented needles. The wavelength of the incident unpolarized light was taken to be 0.65 μm ; at 0.44 μm the polarization curves are similar, although the maximum degree of polarization (for needles and disks) is lower.

We know of no other absorbing materials that, when in the form of small particles, could give the measured 10-fold increase in extinction in going from 0.65 to 0.44 μm . Carbon is ruled out because the imaginary part of its refractive index does not vary much over the visible spectrum. From inspection of measured⁵ optical constants of magnetite (Fe_3O_4), we can rule out this oxide of iron: k does not change sufficiently over the visible spectrum.

The measurements of k for haematite peak at about 0.40 μm , then drop sharply to rise again at $\sim 0.34 \mu\text{m}$. Measurements near these two wavelengths would therefore provide further evidence to consider in deciding if haematite is indeed the material of which the dust particles observed by Ackerman *et al.* are composed.

Neither the observations of Ackerman *et al.* nor the interpretations made here are without precedent. Iron oxide and silicate particles are thought to form as a condensate of meteor ablation; $\sim 1,000$ particles cm^{-3} , of radius $\sim 100 \text{\AA}$, are expected at 60 km (ref. 4). Appreciable horizontal convergence of meteoric material in the 55–75 km region is expected from simple general circulation models⁶. Ionic iron in various oxides and hydrates has been detected at 65 km and above using mass spectrometry⁷, which supports the hypothesis of an extraterrestrial source. Finally, a surface source of haematite particles has been inferred from aircraft observations of transmitted solar radiation over an Asian desert⁸.

We thank Professor Donald Huffman for optical constants of haematite, and Mr Timothy Nevitt for help with some computations.

Received 10 February; accepted 10 May 1984.

1. Ackerman, M., Lippens, C., Muller, C. & Vignault, P. *Nature* **299**, 17–20 (1982).
2. Ackerman, M., Lippens, C., Muller, C. & Vignault, P. *Aeronomica Acta* No. 246 (Belgisch Instituut voor Ruimte Aeronomie, Brussels, 1982).
3. Bohren, C. F. & Huffman, D. R. *Absorption and Scattering of Light by Small Particles* (Wiley, New York, 1983).

4. Hunten, D. M., Turco, R. P. & Toon, O. B. *J. atmos. Sci.* **37**, 1342–1357 (1980).
 5. Huffman, D. R. & Stapp, J. L. in *Interstellar Dust and Related Topics*, (eds Greenberg, J. M. & van de Hulst, H. C.) 291–301 (Reidel, Boston, 1973).
 6. Fiocco, G. & Grams, G. *J. atmos. terr. Phys.* **88**, 512–524 (1971).
 7. Goldberg, R. A. & Witt, G. *J. geophys. Res.* **82**, 2619–2627 (1977).
 8. Kondratyev, K. Ya., Vassilyev, O. B., Grishechkin, V. S. & Ivlev, L. S. *Appl. Opt.* **13**, 478–486 (1974).

Geopotential harmonics of orders 15 and 30

D. G. King-Hele & Doreen M. C. Walker

Royal Aircraft Establishment, Farnborough, Hants GU14 6TD, UK

The Earth's gravitational potential is usually expressed as a double infinite series of spherical harmonics. In recent models^{1–3}, derived from photographic, radio and laser observations of satellites, gravimetry and satellite altimetry, the harmonics have been evaluated up to a degree and an order of 36 or more, so that there are 1,296 or more coefficients to be evaluated. The geoid surfaces derived from such models have errors of approximately 1 m, but much better accuracy is required to take advantage of satellite altimeter measurements accurate to 10 cm or better. So there is a continuing demand to improve the models. The accuracy of the individual coefficients is questionable, and difficult to estimate⁴. The most precise technique for determining coefficients of a particular order is by analysis of satellite orbits which experience resonance with the Earth's gravitational field. We have recently⁵ re-evaluated individual harmonic coefficients of orders 15 and 30 from analyses of 24 orbits having a wide range of inclinations to the Equator. These coefficients have standard deviations equivalent to an accuracy in geoid height of 1 cm, for degree up to 23, and are valuable as a standard against which the comprehensive gravity field models can be tested.

A satellite orbit experiences resonance when the ground track over the Earth repeats after an integral number of revolutions. The most useful resonance is 15th order, when the successive Equator crossings are 24° apart in longitude and the tracks repeat each day after 15 revolutions. This resonance occurs when the average height of the satellite is near 500 km. Many satellites pass through 15th-order resonance as their orbits contract under the action of air drag, and, if the orbits are nearly circular, they may pass through the resonance slowly enough to ensure that the perturbations⁶ due to harmonics of 15th order (and to a lesser extent 30th order) build up to a magnitude that can be measured and analysed⁷. The best satellites of this type, such as 1963-24B, 1971-54A and 1974-34A, have remained near resonance for 2 years or more^{8,9}. Analysis of the changes in orbital inclination and eccentricity of such satellites can yield accurate values for a 'lumped harmonic'—a linear sum of individual harmonic coefficients. When lumped harmonics are available for numerous satellites widely and evenly spread in inclination, it is possible to solve for the individual harmonic coefficients. We have used 24 such analyses, at inclinations between 31° and 144°, to determine the individual 15th-order coefficients of degree 15, 16, ..., 35. Eight of these 24 analyses also give good values for lumped 30th-order harmonics of even degree, which have been successfully used to evaluate individual 30th-order coefficients of degree 30, 32, ..., 40.

To define the harmonic coefficients, we write the longitude-dependent part of the Earth's gravitational potential at an exterior point (r, θ, λ) in normalized form¹⁰ as

$$\frac{\mu}{r} \sum_{l=2}^{\infty} \sum_{m=1}^l \left(\frac{R}{r} \right)^l P_l^m(\cos \theta) \{ \bar{C}_{lm} \cos m\lambda + \bar{S}_{lm} \sin m\lambda \} N_{lm}$$

where r is the distance from the Earth's centre, θ the colatitude, λ the longitude (positive to the east), μ is the gravitational constant for the Earth ($398,600 \text{ km}^3 \text{ s}^{-2}$), R is the Earth's equatorial radius (6,378.1 km), $P_l^m(\cos \theta)$ is the associated

Legendre function of order m and degree l , and $\bar{C}_{lm}$ and $\bar{S}_{lm}$ are the normalized tesseral harmonic coefficients to be evaluated. The normalizing factor N_{lm} is given by

$$N_{lm}^2 = 2(2l+1)(l-m)!/(l+m)!$$

In this notation, we have evaluated the $\bar{C}_{lm}$ and $\bar{S}_{lm}$ for $m=15$ with $l=15, 16, \dots, 35$; and for $m=30$ with $l=30, 32, \dots, 40$. Our method is similar to that described for a previous determination of 15th-order coefficients¹¹, and consists of a simultaneous least-squares solution of (1) the equations for the lumped harmonics, and (2) constraint equations of the form $\bar{C}_{l,m} = 0 \pm 10^{-5} l^{-2}$, and similarly for $\bar{S}_{l,m}$. There were also relaxations in the accuracies of some ill-fitting values of lumped harmonics, so as to ensure that the weighted residuals were <1.4 .

The new results are more reliable than before, because a gap in the coverage of inclination has been filled¹², and also more accurate because of the new data and revisions of previous analyses. Our new values for the 15th-order coefficients, with standard deviations, are given in Table 1, divided into odd and even degree, because the odd-degree values, derived from analysis of inclination, are generally more accurate than the even-degree values, derived from analysis of eccentricity. Also, as expected, the accuracy deteriorates as the degree of coefficients increases: further orbits at low inclinations are required to improve the high-degree values.

For degree 15–23 the values in Table 1 have an average standard deviation of 1.2×10^{-9} , which is equivalent to about 1 cm in the accuracy of the geoid undulation due to 15th-order harmonics. These well-established values are the most useful for testing comprehensive models.

Table 1 Values for 15th-order coefficients

l	$10^9 \bar{C}_{l,15}$	$10^9 \bar{S}_{l,15}$	l	$10^9 \bar{C}_{l,15}$	$10^9 \bar{S}_{l,15}$
15	-20.7 ± 0.5	-6.5 ± 0.4	16	-12.1 ± 2.3	-21.7 ± 1.5
17	7.3 ± 0.8	2.4 ± 0.7	18	-42.4 ± 1.7	-22.3 ± 1.1
19	-16.2 ± 0.7	-13.7 ± 0.6	20	-23.5 ± 2.0	-6.0 ± 1.5
21	17.9 ± 0.6	10.8 ± 0.9	22	23.9 ± 2.0	10.2 ± 1.6
23	20.6 ± 1.3	2.0 ± 1.3	24	0.4 ± 3.6	-22.1 ± 3.2
25	-6.0 ± 1.8	1.1 ± 2.1	26	-14.3 ± 5.5	14.3 ± 5.3
27	-4.6 ± 1.3	9.8 ± 2.4	28	-15.2 ± 6.3	-8.4 ± 6.3
29	-6.9 ± 1.4	-4.0 ± 1.4	30	-3.1 ± 6.7	-15.7 ± 6.2
31	18.4 ± 2.4	-4.9 ± 3.4	32	9.2 ± 5.9	3.0 ± 5.0
33	-1.1 ± 2.8	-12.4 ± 3.6	34	10.4 ± 5.8	5.3 ± 4.9
35	-10.5 ± 4.0	4.2 ± 4.4			

Table 2 Comparison of our values with GEM 10B

l	$10^9 \bar{C}_{l,15}$		$10^9 \bar{S}_{l,15}$	
	Table 1	GEM 10B	Table 1	GEM 10B
15	-20.7	-19.7	-6.5	-6.4
16	-12.1	-14.4	-21.7	-27.8
17	7.3	2.5	2.4	4.8
18	-42.4	-48.3	-22.3	-18.6
19	-16.2	-20.6	-13.7	-15.3
20	-23.5	-23.9	-6.0	4.8
21	17.9	16.2	10.8	9.5
22	23.9	24.1	10.2	-1.3
23	20.6	15.4	2.0	4.1

Table 3 Four-coefficient solution for even-degree 30th-order harmonics, with GEM 10B for comparison

l	$10^9 \bar{C}_{l,30}$		$10^9 \bar{S}_{l,30}$	
	Our values	GEM 10B	Our values	GEM 10B
30	-2.6 ± 1.0	-5.2	8.6 ± 0.6	11.1
32	-8.3 ± 2.7	-0.6	-3.3 ± 2.1	-0.2
34	-12.8 ± 3.2	-11.9	8.8 ± 2.5	1.2
36	-8.9 ± 3.2	-3.9	-1.3 ± 2.6	-0.9

Previous comparisons of gravity field models¹³ suggest that only the recent models are good enough for testing; of these, the Goddard Earth Model 10B (ref. 1) is believed to be quite independent of our values, but the 1981 model of Rapp² and the European model GRIM 3 (ref. 3) use our earlier values and therefore do not provide a useful comparison. Table 2 gives our values for degree 15–23 and those from GEM 10B.

The mean numerical difference between our values and GEM 10B in Table 2 is 3.6×10^{-9} . As the mean standard deviation of our values is 1.2×10^{-9} , this crude but effective comparison suggests that, for order 15 and degree 15–23, the accuracy of GEM 10B is about 3 or 4×10^{-9} , much better than is indicated by other evidence⁴.

For the 30th-order coefficients, we have previously given only a tentative solution¹⁴. Our new data are much better than before, and there is a 40% improvement in the standard deviations of the first three pairs of coefficients when we solve for six coefficients (degree 30, 32, . . . , 40). The average standard deviation of the coefficients of degree 30, 32 and 34 is 2.0×10^{-9} , which is equivalent to an uncertainty of 1.5 cm in the 30th-order contribution to the geoid profile. We also have a satisfactory solution for only four coefficients, up to degree 36, and GEM

10B, which ceases at degree 36, can most appropriately be compared with this four-coefficient solution (see Table 3).

The mean numerical difference between the eight GEM values and ours is 3.7×10^{-9} , while our mean standard deviation is 2.2×10^{-9} . For 30th order, Rapp's model uses our previous values for the (30, 30) coefficients: for degree 32, 34, . . . , 40, his values differ from those of our six-coefficient solution by 4.2×10^{-9} on average. GRIM 3 does not use our previous values, and, in the GRIM3-L1 version¹⁵, the mean difference from our values in Table 3 is 4.3×10^{-9} . These comparisons are encouraging, because the errors in all three models could be as low as 3×10^{-9} , smaller than would be expected from statistical tests⁴. However, as the mean numerical value of the 16 coefficients in Table 3 is only 5.6×10^{-9} , it could also be said that the values in the three models do not much exceed their likely errors.

The accuracy of these high-order coefficients clearly needs to be further improved, but Rapp's study² suggests that the main errors in geoid height arise from errors in the harmonics of order between 3 and 10. The values of these harmonics could, in principle, also be improved by analysing resonances of order 3–10; in practice, however, there are very few satellites in orbits suitable for such analyses.

Received 19 March; accepted 24 May 1984.

1. Lerch, F. J., Putney, B. H., Wagner, C. A. & Klosko, S. M. *Mar. Geod.* **5**, 145–187 (1981).
2. Rapp, R. H. Rep. No. 322 Ohio State Univ. (1981).
3. Reigber, C., Balmino, G., Moynot, B. & Mueller, H. *Manuser. Geodaet.* **8**, 93–138 (1983).
4. Lambeck, K. & Coleman, R. *Geophys. J. R. astr. Soc.* **74**, 25–54 (1983).
5. King-Hele, D. G. & Walker, D. M. C. Tech. Rep. 84021, Royal Aircraft Establishment (1984).
6. Allan, R. R. *Planet Space Sci.* **15**, 53–76 (1967).
7. Gooding, R. H. *Nature phys. Sci.* **231**, 168–9 (1971).

8. Walker, D. M. C. *Geophys. J. R. astr. Soc.* **67**, 1–18 (1981).
9. King-Hele, D. G. *Proc. R. Soc. A* **375**, 327–350 (1981).
10. Kaula, W. M. *Theory of Satellite Geodesy* (Blaisdell, Waltham, 1966).
11. King-Hele, D. G. & Walker, D. M. C. *Proc. R. Soc. A* **379**, 247–288 (1982).
12. Walker, D. M. C. Tech. Rep. 84008, Royal Aircraft Establishment (1984).
13. Klokocnik, J. & Pospisilova, L. *Planet. Space Sci.* **29**, 653–671 (1981).
14. King-Hele, D. G. & Walker, D. M. C. *Planet. Space Sci.* **30**, 411–425 (1982).
15. Reigber, C., Müller, H., Bosch, W., Balmino, G. & Moynot, B. *J. geophys. Res.* (in the press).

A crustal conductor on Eyre Peninsula, South Australia

Antony White & P. R. Milligan

School of Earth Sciences, The Flinders University of South Australia, Bedford Park, South Australia 5042, Australia

Short period variations in the geomagnetic field components were recorded in a closely spaced network of sites on southern Eyre Peninsula, South Australia. They reveal the existence of a linear zone of very good electrical conductivity within the late Archaean–early Proterozoic Gawler Craton. There is strong enhancement of the horizontal component variations transverse to the zone and an abrupt phase change in the vertical component, particularly for periods near 2×10^3 s. These features are consistent with a line or narrow thin sheet of current flowing in a conductor at a depth ≤ 10 km; this almost certainly requires a major fracture or shear zone within the crystalline basement, for which there is supporting geophysical evidence. Superficial Cenozoic deposits blanket much of the area and consequently the basement geology is not known in detail, although there are structures parallel to the conductor 20 km to the east. Identification of a major linear crustal feature in such an area demonstrates the usefulness of geomagnetic variation studies for understanding deep geology.

The feature was discovered during a study of the coast effect in South Australia^{1,2}. In this effect the flow of induced electrical current in the highly conductive ocean produces large vertical magnetic field variations at the coast, and these are strongly correlated with the horizontal field variations perpendicular to the coast^{3,4}. A coastal profile in the south-east of South Australia showed a normal response pattern¹ but unpublished data from profiles on Eyre Peninsula showed strong disturbance of the coast effect. This prompted a detailed study with some of the stations only 5 km apart.

Figure 1 shows the magnetometer sites and the regional geology of the area⁵. Magnetometers were generally deployed along NW–SE profiles; Fig. 2 shows magnetograms from the profile closest to the coast. Variations in the north horizontal

component (H) do not differ discernibly from station to station; indeed, this is true over the whole study area. The east horizontal component (D) variations, while very similar in form from station to station, show amplitude enhancement by as much as a factor of 2 for station CB with respect to station PL. The enhancement is frequency-dependent, with a maximum effect at periods of $\sim 2 \times 10^3$ s, but with almost no effect at the diurnal period. In the vertical component (Z) the short period variations are strongly attenuated at station CB and there is a large phase change between stations to the north-west and those to the south-east.

Similar effects are seen in magnetograms from stations along the more northerly profiles, with a peak in the D component enhancement and a 'null' in the Z component occurring close to (but probably just east of) a line joining stations CB, BL, CU and BR. It is useful to compute transfer functions between Z (assumed to be caused wholly by induced telluric currents) and the normal, non-anomalous horizontal field. Previous observations on Eyre Peninsula showed that the D component at PL is not enhanced, so that the horizontal field at this station was taken to be the reference normal field. Figure 3 shows transfer functions for most stations in a vector form⁴ for a 3,600-s period. The direction of the in-phase arrows has been reversed so that they point towards, rather than away from, concentrations of telluric current, and their length represents the ratio of vertical to horizontal field variations. The coast effect is clearly being severely disrupted by what seems to be a relatively narrow and shallow line current flowing in a zone whose axis is indicated in Fig. 3. This is probably a case of current being channelled from the ocean into a zone whose total conductance compares favourably with that of the seawater layer. Modelled as a simple buried line current, the maximum depth is estimated as 15 km, although it is probably considerably less.

The remarkable aspect of this conducting zone is that in neither the relatively thin surface cover nor in the interpreted basement geology, is there any real indication of a coincident structure with which the zone might be associated. To the west are limited exposures of the late Archaean Sleaford complex partly overlain by basal units of the Hutchison Group metasediments. To the east there is better exposure of the Hutchison and the Lincoln complex. The latter is thought to be composed of

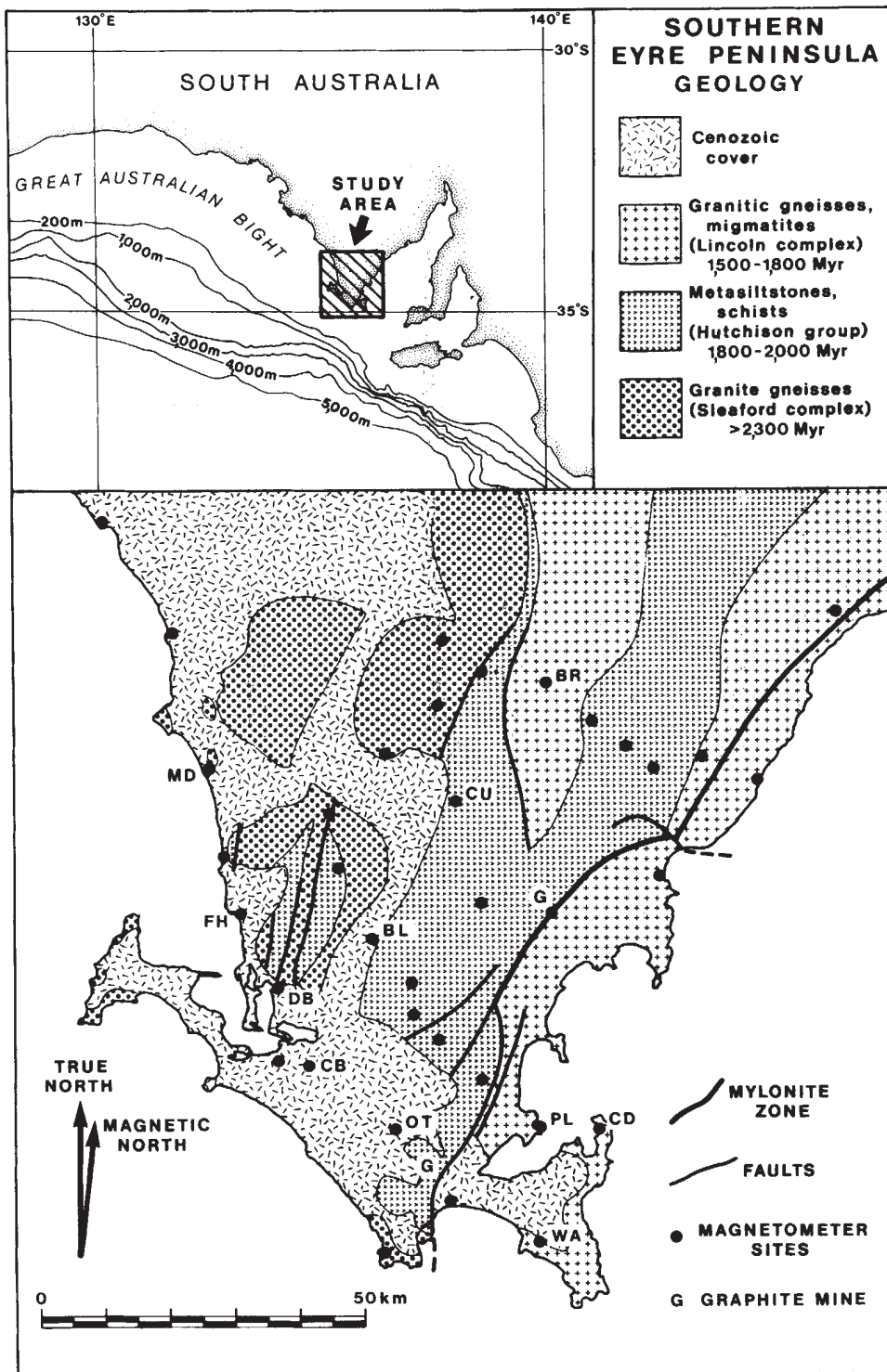


Fig. 1 Study area location showing magnetometer sites and general geology. Sites referred to in the text and Fig. 2 are identified. (After ref. 5).

granite intrusives and basement reworked by the Kimban orogeny (~1,820–1,580 Myr) which also caused intense deformation to the Hutchison Group⁶. This part of the Gawler Craton which was severely affected by Kimban tectonism, is termed the Cleve sub-domain. The adjacent Coultas sub-domain is a major cratonic area comprising rocks of the Sleaford complex, and was much less affected by the orogeny^{5,6}. It seems probable that the conducting zone identifies the presently ill-defined boundary between these two sub-domains.

There is other geophysical evidence to support the idea of a major fracture or shear zone in the crust. There are several static magnetic anomalies closely parallel to the major structural

trends east of the conducting zone, but almost none to the west⁷. The Bouguer gravity⁸ and seismicity (Fig. 4) also show some correlation with the zone. Although the seismicity is relatively low, the earthquake epicentres are concentrated closer to the axis of the conducting zone than to the mapped faults. All of the earthquakes originate within the crust, most within the top 12 km. The mylonite zone, although of Kimban origin, was reactivated in the Tertiary when the block faulting commenced that caused the graben structures of the South Australian gulfs. It is significant that the present seismicity is not closely associated with the mylonite zone but with an unidentified feature 20 km to the west. The Bouguer gravity displays a prominent

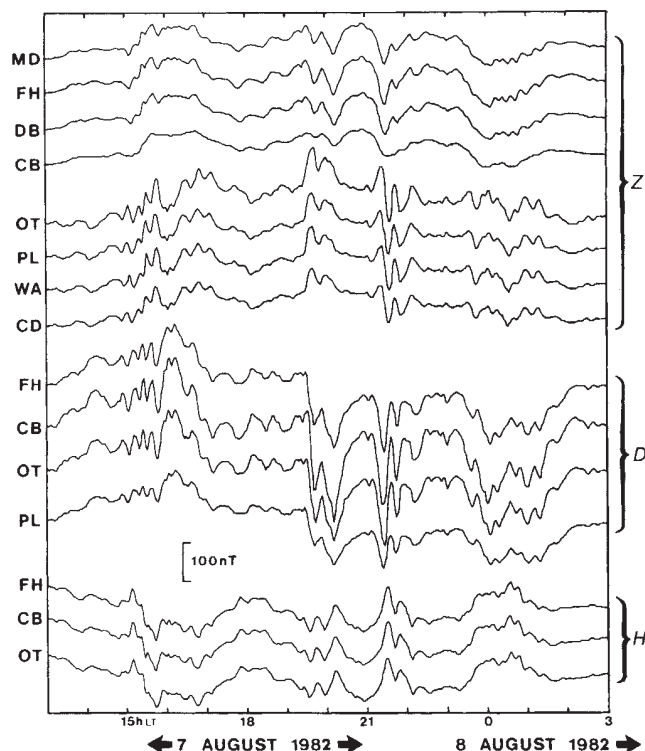


Fig. 2 Simultaneous magnetograms from selected stations (identified in Fig. 1). The traces show the small variations occurring in the strengths of the horizontal north (*H*), horizontal east (*D*) and vertical (*Z*) components of the geomagnetic field. Note the large change in phase and amplitude of the *Z*-component variations for stations above and below stations CB.

fault-like gradient coincident with the southern half of the conducting zone. A simple analysis of this as a vertical fault gives a depth estimate of 6–8 km to the centre of the fault. The preliminary interpretation, therefore, is that the conductor lies within a major fracture or shear zone within the top 10–15 km of the crust.

The conductive mechanism necessary to produce the geomagnetic variation anomaly is unknown. Numerous water bores, particularly in the southern part of the peninsula, show the cover to be <100 m thick and therefore localized conduction within surface sediments cannot explain the magnitude of the anomaly. Other workers have reported such shallow highly conducting zones which also produce enhancement of the horizontal field^{9–16}. The North American Central Plains anomaly^{9,10} is one such comparison that has been attributed to graphitic schists within a Precambrian metamorphic belt and possibly associated with a Proterozoic plate boundary¹¹. A graphitic schist near the base of the Hutchison Group outcrops a few kilometres west of the mylonite zone, and has been mined at two locations (Fig. 1). However, we detect little significant effect on the geomagnetic variations in this region. The reappearance of this unit in the vicinity of the conducting zone is possible but it has not been reported, and it would apparently require more extensive graphitization to cause the observed effect. Nevertheless, a marked increase in conductivity is caused by even small amounts of graphite¹⁷, whether within a metasedimentary schist or a shear zone, and it remains a possible candidate. The most probable mechanism is the occurrence of saline waters within a fracture or shear zone¹⁸. The implications are that such a feature probably would have to be several kilometres wide and extend deep into the crust. This is quite plausible in the overall geological context and its discovery emphasizes the potential of the geomagnetic variation method in probing crystal geology.

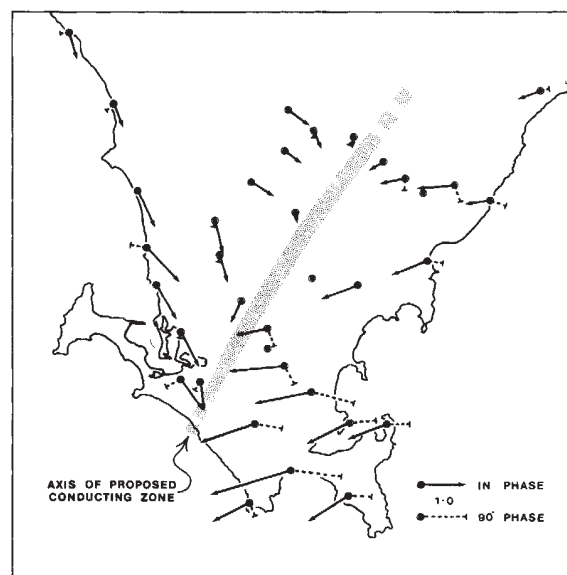


Fig. 3 Transfer functions relating the vertical to the horizontal field variations for a 3,600-s period. The direction is that of the horizontal field showing highest correlation with the vertical field. In-phase arrows have been reversed.

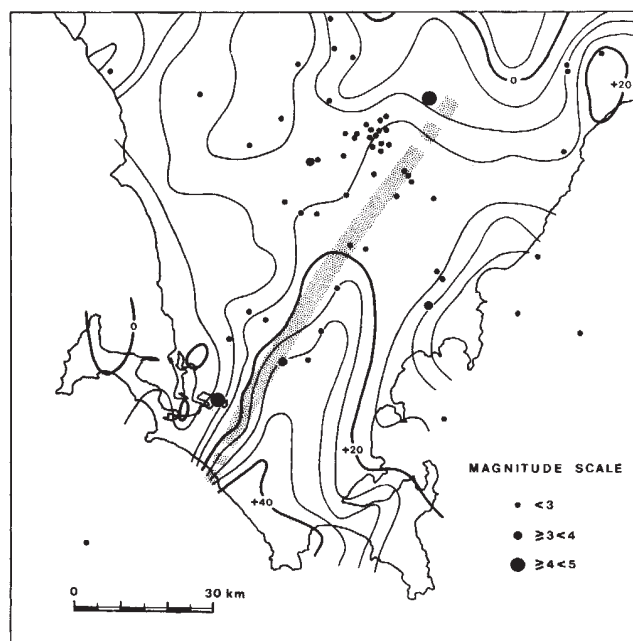


Fig. 4 Bouguer gravity and seismicity data for southern Eyre Peninsula. All but three epicentres are located instrumentally. About half of the locations have an uncertainty of ± 3 km, the remainder ± 10 km. Gravity contour interval, 5 mGal.

We thank the Australian Research Grants Scheme and Flinders University Research Committee for financial support. We have had helpful discussions with A. J. Parker and R. B. Flint of the South Australian Department of Mines and Energy, and the C.R.A. personnel on Eyre Peninsula. Bob Nation of the Adelaide University Seismology Unit supplied the seismicity data.

Received 16 December 1983; accepted 9 May 1984.

1. White, A. & Polatajko, O. *J. Geomagn. Geoelect.* **30**, 109–120 (1978).
2. White, A. & Hopgood, D. N. *J. Geomagn. Geoelect.* **31**, 479–484 (1979).
3. Parkinson, W. D. *Geophys. J. R. astr. Soc.* **2**, 1–14 (1959).
4. Schmucker, U., *Bull. Scripps Inst. Oceanogr.* **13**, 1–165 (1970).
5. Thomson, B. P. (compiler) *Geological map of South Australia* (Department of Mines and Energy, Adelaide, 1980).

6. Parker, A. J. & Lemon, N. M. *J. geol. Soc. Aust.* **29**, 221–238 (1982).
7. *Total Magnetic Intensity map South Australia 1:2,000,000* (Department of Mines and Energy, Adelaide 1982).
8. Gerdes, R. A. (compiler) *Lincoln Bouguer anomaly map* (Department of Mines and Energy, Adelaide 1973).
9. Camfield, P. A., Gough, D. I. & Porath, H. *Geophys. J. R. astr. Soc.* **22**, 201–221 (1971).
10. Alabi, A. O., Camfield, P. A. & Gough, D. I. *Geophys. J. R. astr. Soc.* **43**, 815–833 (1975).
11. Camfield, P. A. & Gough, D. I. *Can. J. Earth Sci.* **14**, 1229–1238 (1977).
12. Niblett, E. R., De Laurier, J. M., Law, L. K. & Plet, F. C. *J. Geomagn. Geoelect.* **26**, 203–221 (1974).
13. Wilhelm, J. & Friis-Christensen, E. *J. Geomagn. Geoelect.* **26**, 173–189 (1974).
14. Babour, K. *et al. Geophys. J. R. astr. Soc.* **45**, 538–600 (1976).
15. de Beer, J. H. & Gough, D. I. *Geophys. J. R. astr. Soc.* **63**, 479–495 (1980).
16. Nityananda, N. & Jayakumar, D. *Phys. Earth planet. Inter.* **27**, 223–228 (1981).
17. Shankland, T. J. & Duba, A. G. *EOS* **63**, 438 (1982).
18. Shankland, R. J. & Ander, M. E. *J. geophys. Res.* **88**, 9475–9484 (1983).

Mantle-derived Archaean monzodiorites and trachyandesites

Steven B. Shirey* & Gilbert N. Hanson

Earth and Space Sciences, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

The 2,680–2,750-Myr old^{1–5} intrusive granite–greenstone terranes^{6–8} in the Rainy Lake region, Ontario, and in, and adjacent to, the Vermilion District, Minnesota, of the Superior Province in North America, include a wide variety of volcanic and intrusive rocks with initial Pb, Sr and Nd isotope ratios close to those for the mantle^{4,5,9–12}. The isotopic constraints require the production of large volumes of silica-oversaturated magmas from silica-undersaturated mantle within 100–200 Myr. We present here geochemical data on monzodiorites and trachyandesites from the Rainy Lake area which are strongly enriched in large-ion-lithophile elements (LILE). We conclude that they are derived by direct melting of the mantle at depths of <50 km in either anhydrous or hydrous conditions. Their rare-earth element (REE) abundances and initial ¹⁴³Nd/¹⁴⁴Nd ratios suggest that their mantle sources were enriched in LILE shortly before melting. The monzodiorites–trachyandesites and their granodioritic derivatives may comprise up to 20% of the exposed igneous rocks in the Rainy Lake region and Vermilion District. If these rocks are as abundant in other Archaean terranes, a significant part of the early continental crust could have formed by direct melting of LILE-enriched mantle.

Rocks occurring as monzodiorites and andesites through trachyandesites in the Rainy Lake region form a suite of cogenetic rocks. They occur as volcanic units in the Keewatin Series, as intrusive and volcanic units in the Rocky Islet Bay complex and as the major component of the Couthiching metasedimentary rocks. The more primitive members of the sequence are silica-oversaturated with normative orthopyroxene usually greater than normative clinopyroxene, normative plagioclase about An₅₀ or less, Mg numbers of >0.60, high Ni and Cr contents, strong light rare earth element (REE) enrichment, heavy REE abundances ~6–12 times that in chondrites, and relatively high K, Sr, Zr and Nb concentrations (Table 1, Fig. 1). They could not have originated by contamination of basic magma by felsic rocks because they have higher Ce/Yb ratios and abundances than the exposed crustal rocks that could have been the contaminants.

This suite of rocks has major- and trace-element characteristics similar to the Miocene high-Mg andesites that are termed sanukites in the Setouchi volcanic belt, Japan^{13,14}, and to high-Mg andesites from the Archaean Welcome Well volcanic complex, Australia^{15,16}. The high-Mg andesites from both areas have been explained as being essentially primitive magmas derived by direct partial melting of the mantle^{13–16}. Similar magma types occur in other intrusive granite–greenstone terranes throughout

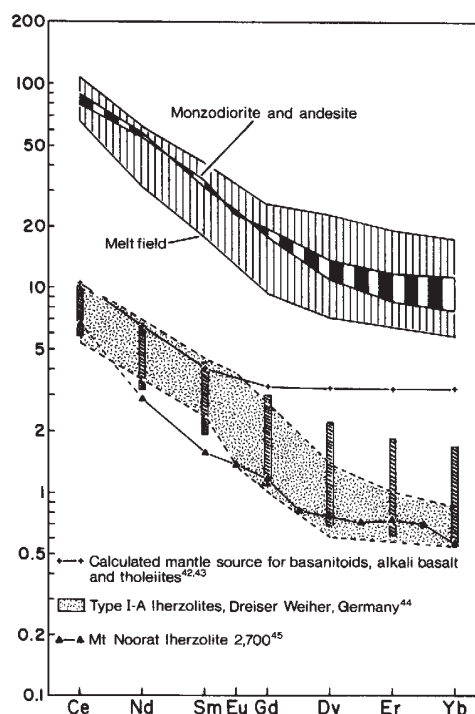


Fig. 1 Chondrite normalized rare earth element (REE) diagram showing: field of typical primitive monzodiorite and andesite from the Rainy Lake region (samples 5–80 and 53–80 in Table 1); calculated light REE-enriched mantle sources for alkali basalts, basanoids and tholeiites^{42,43}; pattern and field for LILE-enriched lherzolites^{44,45}; estimated typical range for light-REE enriched mantle sources shown by vertical bars; and field for melts derived by 8% batch melting of the typical range for light-REE enriched mantle sources leaving 60% olivine, 30% orthopyroxene and 10% clinopyroxene in the residue.

the world (Table 1). Their abundance, however, and that of their derivatives, is unknown.

Sanukites are glassy andesites with olivine, clinopyroxene or orthopyroxene phenocrysts. Koto¹⁷ used the term 'sanukitoid' to describe all textural modifications of consolidated sanukite magmas¹⁸. We recommend that Archaean plutonic and volcanic rocks with the unique geochemical characteristics of sanukites, that is, silica-oversaturated melts with high Mg numbers, high Ni, Cr and LILE abundances, be considered as members of the sanukitoid suite.

The most primitive sanukitoids from the Rainy Lake region and other greenstone belts plot close to the melt field for pyrolite on the olivine saturation surface in Fig. 2. Their high Ni contents (180–200 p.p.m.) compared with their Mg numbers (0.60–0.64) suggest that they have not undergone extensive fractional crystallization and that they are derived from primary melts lying close to or in the melt field just above the 1,200 °C isotherm for 1-bar pressure in Fig. 2. The field of the most primitive Archaean sanukitoids in Fig. 2 is nearly identical to the fields for the Setouchi volcanic belt, Japan¹⁴, and for experimentally derived silica-oversaturated liquids formed by melting in both hydrous and anhydrous conditions at low pressures^{19,20}.

Experimental studies are consistent with a direct mantle derivation for silica-oversaturated melts at pressures of <5 kbar and at temperatures of ~1,150–1,250 °C in anhydrous conditions^{19,21,22}. Melting in hydrous conditions results in silica-oversaturated melts at higher pressures²³. The common occurrence of amphibole in Archaean sanukitoid rocks suggests that their parent melts may have been hydrous, although water may have been introduced at crustal levels.

Figure 1 and Table 1 show the light REE, Sr, Zr and Nb enrichment characteristic of the Archaean monzodiorites and trachyandesites. Compared with the Miocene sanukitoids, the

* Present address: Carnegie Institution of Washington, Department of Terrestrial Magnetism, 5241 Broad Branch Road, NW, Washington, DC 20015, USA.

Table 1 Comparison of andesites, trachyandesites and monzodiorites from various Archaean greenstone belts with a Setouchi high-Mg andesite and a melt derived by anhydrous partial melting of a parent of pyrolite composition at 2kbar

	Upper Archaean crust A	Rainy Lake Ontario 5-80 B	Rainy Lake Ontario 53-80 C	Rainy Lake Ontario 41-66 D	Vermilion district Minnesota B-18 E	Jack- Lake Ontario F126 F	Suomus- salmi Finland SU10 G	Welcome Well Australia W131 H	Marda Complex Australia I I	Que Que Zimbabwe Rh73155 J	Setou chi Japan TK-52 K	Jaques Green 1,200 °C 2kbar L
SiO ₂	57.4	54.55	55.93	58.0	54.6	56.35	52.90	53.01	55.27	57.62	56.40	53.0
TiO ₂	0.9	0.72	0.81	0.71	0.73	0.67	0.73	0.76	0.27	0.56	0.64	2.6
Al ₂ O ₃	15.6	14.38	14.78	14.3	15.1	15.93	13.11	17.70	14.71	15.97	15.60	13.5
Fe ₂ O ₃	—	8.56	8.73	2.16	3.04	—	8.35	7.28	1.95	1.33	3.20	0.5
FeO	9.5	—	—	5.02	4.40	6.74	—	—	6.41	4.68	2.88	7.3
MnO	—	0.12	0.14	0.11	0.14	0.12	0.16	0.10	0.14	0.08	0.11	0.1
MgO	5.2	7.98	6.72	5.99	6.40	6.29	9.47	5.99	6.06	6.23	8.35	10.5
CaO	7.3	8.71	7.57	5.96	7.42	7.97	8.82	9.64	7.78	7.41	6.38	9.7
Na ₂ O	3.1	2.87	3.38	3.61	4.27	3.82	1.63	4.92	2.94	3.51	2.99	2.3
K ₂ O	0.9	1.34	1.73	2.43	2.22	1.74	2.15	0.49	1.21	0.43	1.73	0.5
P ₂ O ₅	—	0.36	0.30	0.35	0.34	0.37	0.46	0.36	0.17	0.1	0.17	—
Total	99.9	99.59	100.09	98.64	98.66	100.00	97.78	100.25	96.91	97.92	98.45	100.00
Cr	140	469	429	291	—	—	770	67	299	156	488	—
Ni	90	188	163	159	—	108	140	147	205	105	181	—
Sr	300	627	650	1164	1237	1413	390	1350	293	—	—	—
Zr	100	138	167	140	—	53	115	115	135	—	—	—
Nb	5	7.0	9.2	7.3	—	2	—	5.9	—	—	—	—
Ce	26.8	64.7	72.7	86.4	94.6	69	83.7	63	57	24.6	35.7	—
Nd	13.0	33.4	34.7	42.4	44.0	—	44.5	33	22	12.0	—	—
Sm	2.78	6.39	6.20	7.23	8.15	—	8.13	—	3.7	2.54	3.22	—
Eu	0.9	1.64	1.77	1.77	2.12	—	2.16	—	1.14	0.81	0.94	—
Gd	2.85	5.10	3.36	4.85	6.37	—	6.10	—	3.7	2.23	—	—
Dy	2.93	4.44	3.51	3.47	4.04	—	3.85	—	2.8	1.95	—	—
Er	1.81	2.50	1.80	1.79	1.97	—	1.85	—	1.69	1.18	—	—
Yb	1.79	2.31	1.61	1.64	1.92	—	1.65	—	1.57	1.11	1.89	—
Mg no.	—	0.65	0.60	0.61	0.62	0.62	0.69	0.62	0.57	0.65	0.72	0.71
Liquidus temperature (°C)	—	1,238	1,223	1,240	1,235	1,213	1,269	1,294	1,195	1,211	1,286	1,284
Equil. Ol.	—	85.4	83.0	83.5	84.6	84.6	87.6	80.3	80.3	85.7	89.4	87.5

Mg numbers are calculated using total Fe as Fe²⁺. Equilibrium olivine compositions were calculated by the method of Langmuir and Hanson³⁵ using the relation for olivine-melt K_d given by Ford *et al.*³⁶ for anhydrous experiment at 1 atmosphere.

A, Average Archaean upper crust³⁷. B, Monzodiorite, Rocky Islet Bay Complex, Sandpoint I, Rainy Lake. C, Medium-K, basic, calc-alkaline andesite, Keewatin volcanics, Rainy Lake. D, Monzodiorite, Ottentail Lake pluton, Rainy Lake. E, Trachyandesite of Newton Lake Fm, Table 2 of ref. 31. F, Diorite dyke, Jackfish Lake plutonic complex, Wabigoon belt²⁹, Table 4a of ref. 30. G, Fine-grained alkali basalt sample S63³⁸, Table 2 of ref. 38. H, Porphyritic basalt-andesite¹⁶, Table 1 of ref. 16. I, Fine-grained basaltic andesite. Major element analysis, Table II of ref. 39; trace elements³⁹, Table 1 of ref. 40. J, Andesite of Maliyama Fm, Table 2A of ref. 41. K, High-Mg andesite from Setouchi volcanic zone, SW Japan, Table 1 of ref. 14. L, Experimental melt of parent of pyrolite composition, Table 4 of ref. 19.

Archaean suites are more light REE enriched, but both have variability in their light REE enrichment and overall REE abundances.

Both the extent of enrichment and the relative enrichment of the various LILE in the sanukitoids must be explained. If the mantle source had a flat REE pattern and olivine, orthopyroxene and clinopyroxene were the principal phases in the residue, the light-REE enriched patterns for these rocks could have been produced only if the extent of melting were $< \sim 0.2\%$ (ref. 24). Slightly higher extents of melting would have produced melts with similarly high light REE abundances, but the melts would not have had the proper relative enrichments. The presence of amphibole or garnet in the residue would help to achieve the appropriate relative enrichments at higher percentages of melting, but the high abundance of light REE require equally low extents of melting if the mantle had a flat REE pattern. Such small fractions of melt may have difficulty leaving their mantle source²⁵⁻²⁷.

The mantle source for the sanukitoids was probably enriched in LILE. However, the initial ¹⁴³Nd/¹⁴⁴Nd ratios²⁸ indicate that this enrichment could not have occurred more than ~ 100 – 200 Myr before melting. The enrichment of the source could be due to the injection of a small fraction of LILE-enriched magma or metasomatic fluids just before melting.

Figure 1 includes REE patterns for LILE-enriched mantle based on proposed mantle sources for alkali basalts, basanites and tholeiites and actual REE patterns for lherzolites enriched in LILE by metasomatism. Between 5 and 10% melting of such mantle, leaving olivine, orthopyroxene and clinopyroxene in the residue, leads to melts with REE patterns similar to that of the Archaean sanukitoids.

In this region of the Superior Province, the more primitive sanukitoids are associated with greater volumes^{6,8} of evolved quartz monzodiorites, granodiorites and monzogranites: for

example, the Ottentail Lake pluton²⁸ and Jackfish Lake plutonic complex^{29,30} in the Rainy Lake region and the Farm Lake facies of the Giants Range granite³¹ in northeastern Minnesota. Major and trace element chemistry suggest they may be differentiates or partial melts of sanukitoid parents in which clinopyroxene and amphibole would be early minerals to crystallize or would be residual minerals during melting. These phases are silica-undersaturated relative to their parents and experimental evidence³²⁻³⁴ shows that they are especially effective for driving derivative liquids to greater silica oversaturation.

Miocene sanukitoids seem to be restricted to island arc regimes where the melting is presumably hydrous, the water being derived from the subducting slab. While the Archaean melts may have developed in island arc settings, it is not required, because sanukitoid melts can form in both hydrous and anhydrous conditions. Production of silica-oversaturated, mantle-derived melts might have been more common in Archaean. The presumably higher heat content of the mantle might have allowed melting of the mantle at shallow depths under a range of tectonic settings. Even in the same terrane, the melts could have had similar trace and major element characteristics with variable degrees of silica oversaturation or undersaturation depending on the conditions of melting. If water were present during melting, the melts would have been more silica oversaturated. If the melting were dry or in the presence of carbon dioxide, the melts would have been more silica undersaturated. If the melting were deeper, this would favour less silica oversaturation, or silica undersaturation.

The abundance of sanukitoid rocks and their derivatives in other Archaean terranes is unclear. Because they have composition similar to that of the average upper continental crust (Table 1), potentially a significant part of the early continental crust could have been derived directly by shallow melting of LILE-enriched mantle.

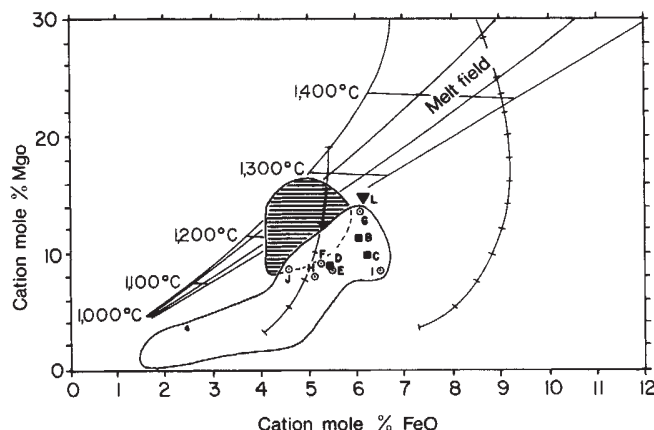


Fig. 2 MgO-FeO diagram for Archean sanukitoid rocks compared with the melt field for pyrolite⁴⁶ at 1 atm with liquid lines of descent for olivine fractionation marked at 5% intervals^{47,48}. The shaded field encloses the Setouchi volcanic zone sanukitoids¹⁴. The most primary sanukitoids from the Rainy Lake region (solid symbols) and the other Archean terranes given in Table 1 plot in the plain field. ▽, The experimental run of Jaques and Green¹⁹ given in Table 1. Letters near the points match the analyses in Table 1. The lower left part of the plain field encloses more evolved rocks from the Rainy Lake region and other areas in the Superior Province^{29,31}.

We thank S. S. Goldich for providing samples for analysis and for field assistance, J. M. Rhodes and J. W. Sparks (University of Massachusetts for facilities for XRF, L. D. Ashwal and J. L. Wooden for providing Nd isotope data on anorthosites and gabbros in advance of publication and J. L. Banner and T. D. Bullen for fruitful discussions. The manuscript benefited from the reviews of E. Krogstad, R. Walker, W. Sharp, K. Spencer and J. Banner. This study was supported by a Sigma-Xi grant-in-aid of research and a GSA Penrose grant to S. B. S. and by NSF grant EAR 7926363 and NASA grant NAGW-486 to G.N.H.

Received 29 December 1983, accepted 16 May 1984.

- Davis, D. W. & Edwards, G. R. *Can. J. Earth Sci.* **19**, 1234-1245 (1982).
- Davis, D. W., Blackburn, C. E. & Krogh, T. E. *Can. J. Earth Sci.* **19**, 254-266 (1982).
- Goldich, S. S. & Peterman, Z. E. *Precamb. Res.* **11**, 307-327 (1980).
- Hanson, G. N., Goldich, S. S., Arth, J. G. & Yardley, D. H. *Can. J. Earth Sci.* **8**, 1110-1124 (1971).
- Jahn, B. M. & Murthy, V. R. *Geochim. cosmochim. Acta* **39**, 1679-1689 (1975).
- Blackburn, C. E. *Kenora-Fort Frances Geological Compilation Series*, Map 2443 (Ontario Geological Survey, Toronto, 1980).
- Goodwin, A. M., in *Proc. 1978 Archean Geochemistry Conf.* (eds Smith, I. E. M. & Williams, J. G.) 73-106 (University of Toronto Press, 1978).
- Sims, P. K. & Morey, G. B. (eds) *Geology of Minnesota: A Centennial Volume* (Minnesota Geological Survey, St Paul, 1972).
- Hart, S. R. & Davis, G. L. *Bull. geol. Soc. Am.* **80**, 595-616 (1969).
- Peterman, Z. E., Goldich, S. S., Hedge, C. E. & Yardley, D. H. *Geol. Soc. Am. Mem.* **153**, 193-217 (1972).
- Tilton, G. R. *Geochim. cosmochim. Acta* **47**, 1191-1197 (1983).
- Shirey, S. B. & Hanson, G. N. *EOS* **64**, 330 (1983).
- Ishizaka, K. & Carlson, R. W. *Earth planet. Sci. Lett.* **64**, 317-340 (1983).
- Tatsumi, Y. & Ishizaka, K. *Earth planet. Sci. Lett.* **60**, 293-304 (1982).
- Giles, C. W. *Spec. Publ. geol. Soc. Aust.* **7**, 275-286 (1981).
- Giles, C. W. & Hallberg, J. A. *Contr. Miner. Petrol.* **73**, 287-310 (1982).
- Koto, B. *J. geol. Soc., Tokyo* **23**, 95-127 (1916).
- Tatsumi, Y. & Ishizaka, K. *Earth planet. Sci. Lett.* **53**, 124-130 (1981).
- Jaques, A. L. & Green, D. H. *Contr. Miner. Petrol.* **73**, 287-310 (1980).
- Green, D. H. *Can. Miner.* **14**, 255-268 (1976).
- Mysen, B. O. & Kushiro, T. *Am. Miner.* **62**, 843-865 (1977).
- Presnall, D. C., Dixon, J. R., O'Donnell, T. H. & Dixon, S. A. *J. Petrol.* **20**, 3-35 (1979).
- Tatsumi, Y. *Earth planet. Sci. Lett.* **60**, 305-317 (1982).
- Hanson, G. N. *Rev. Earth planet. Sci.* **8**, 371-406 (1980).
- Ahern, J. L. & Turcotte, D. L. *Earth planet. Sci. Lett.* **45**, 115-122 (1979).
- Maaloe, S. *Geochim. cosmochim. Acta* **46**, 45-57 (1982).
- Waff, H. S. & Bulau, J. R. *J. geophys. Res.* **84**, 6109-6114 (1979).
- Shirey, S. B. thesis, State Univ. New York, Stony Brook, (1984).
- Longstaffe, F. J., McNutt, R. H. & Crockett, J. H. *Precamb. Res.* **17**, 275-296 (1982).
- Longstaffe, F. J., McNutt, R. H. & Schwarcz, H. P. *Can. J. Earth Sci.* **17**, 1046-1063 (1980).
- Arth, J. G. & Hanson, G. N. *Geochim. cosmochim. Acta* **39**, 325-362 (1975).
- Helz, R. T. *J. Petrol.* **14**, 249-302 (1973).
- Helz, R. T. *J. Petrol.* **17**, 1139-1193 (1976).
- Spulber, S. D. & Rutherford, M. T. *J. Petrol.* **24**, 1-25 (1983).
- Langmuir, C. H. & Hanson, G. N. in *Thermodynamics of Minerals and Melts* (eds Newton, R. C., Navrotsky, A. & Wood, B. J.) 274-271 (Springer, New York, 1981).
- Ford, C. E., Russell, D. G., Craven, J. A. & Fisk, M. R. *J. Petrol.* **24**, 256-265 (1983).
- Taylor, S. R. & McLennan, S. M. *Phil. Trans. R. Soc. A301*, 381-399 (1981).
- Jahn, B. M. *et al. J. Petrol.* **21**, 201-244 (1978).
- Hallberg, J. A., Johnston, C. & Bye, S. M. *Precamb. Res.* **3**, 111-136 (1976).
- Taylor, S. R. & Hallberg, J. A. *Geochim. cosmochim. Acta* **41**, 1125-1129 (1977).
- Hawkesworth, C. J. & O'Nions, R. K. *J. Petrol.* **18**, 487-520 (1977).
- Frey, F. A., Green, D. H. & Roy, S. D. *J. Petrol.* **19**, 463-513 (1978).
- Sun, S. S. & Hanson, G. N. *Contr. Miner. Petrol.* **52**, 77-106 (1975).
- Stosch, H. G. & Seck, H. A. *Geochim. cosmochim. Acta* **44**, 457-470 (1980).
- Frey, F. A. & Green, D. H. *Geochim. cosmochim. Acta* **38**, 1023-1059 (1974).
- Ringwood, A. E. *Composition and Petrology of the Earth's Mantle* (McGraw-Hill, New York, 1975).
- Langmuir, C. H. & Hanson, G. N. *Phil. Trans. R. Soc.* **297**, 383-407 (1980).
- Hanson, G. N. & Langmuir, C. H. *Geochim. cosmochim. Acta* **42**, 725-741 (1978).

Simplifying simple epidemic models

Denis Mollison

Heriot-Watt University, Riccarton, Edinburgh EH14 4AS, UK

Interest has recently revived in the use of simple models for epidemic diseases. In particular, Anderson *et al.*¹ have introduced an improved simple differential equation model for diseases such as fox rabies which regulate the population density of their host. Here I describe how such apparently simple models can be dissected into their basic components. This dissection facilitates a structural sensitivity analysis in which we explore the dependence of features of a model's behaviour on the assumptions regarding each component. I report that particular features can be related to particular components, for example, oscillations depend mainly on population growth and the generation gap of the disease, while estimates of the effect of potential control strategies such as vaccination or culling can be related directly to assumptions concerning the infection term and the way it changes with population density. Some features, for example, the level of prevalence of the disease and the period of any oscillations, turn out to be robust, depending mainly on basic ecological parameters. Others, including the crucial estimates regarding control, prove very sensitive to the details of the model. This is unfortunate, as the detailed form of the components is to a large extent chosen, not for ecological reasons, but to keep models simple, for example, the infectious period of a disease is often assumed to have an exponential distribution because this implies a constant death or recovery rate for infectious individuals which is mathematically very convenient for a continuous-time model. Because our dissection is in terms of components with straightforward ecological interpretations, any required improvements in modelling can be related to observational evidence which either exists or for which experiments can be suggested^{2,3}.

The three basic components essential for modelling the endemic state of a host-regulating disease are: (1) the population growth term, describing the population dynamics in the absence of disease; (2) the infection term, describing the infectious contacts made by a diseased individual; and (3) the 'generation gap' of the disease, that is, the time interval between an individual's becoming infected and his passing on the disease. The model presented by Anderson *et al.*¹ for fox rabies is about as simple as a complete model incorporating all three components can be, but it nevertheless consists of a set of three differential equations in which the relationships between individual components and the model's behaviour are not easy to discern. These equations, slightly simplified, for the densities of susceptible (X), incubating (I) and infectious (Y) individuals, are:

$$\begin{aligned}\frac{\partial X}{\partial t} &= \rho X - \beta XY \\ \frac{\partial I}{\partial t} &= \beta XY - \sigma I \\ \frac{\partial Y}{\partial t} &= \sigma I - \alpha Y\end{aligned}\quad (1)$$

where β , σ and α are constants. In this model, ρX defines the population growth term, βXY the infection term, and σI and αY the distribution of the generation gap. The per capita net growth rate ρ may be density-dependent; usually it is reasonable to assume that it decreases with density, from a value r in low-density populations to zero at the carrying capacity K of the habitat.

A number of such complete models, including a difference equation alternative, with time step equal to the mean generation gap, are analysed in work described elsewhere². Here I shall concentrate on the results and understanding that can be obtained by considering the basic components on their own.

First I shall discuss the infection term in isolation, showing how conclusions on control depend directly on assumptions concerning this term. It is common in simple epidemic models for the infection term to be specified in terms of the overall rate at which infectious contacts occur; for example, the infection rate is commonly assumed to have a multiplicative form, βXY , as in equation (1). However, a better understanding is achieved if we work in terms of the mean number of potentially infectious contacts made by an infectious individual, R_0 (often called the 'basic reproductive rate' of the epidemic, although 'ratio' would be more accurate). In most cases it is reasonable to assume (i) that R_0 does not depend on the density of infectives Y , so that the overall contact rate is $(R_0/\tau_2)Y$, where τ_2 denotes the mean infectious period [$\tau_2 = 1/\alpha$ in the model of equation (1)]. If we also assume (ii) that a proportion X/N of such contacts are successful, as will be the case if a population of density N mixes homogeneously (here $N = X + I + Y$), the mean number of successful infectious contacts (the 'effective reproductive ratio') will be $R = R_0 X/N$. The overall infection rate will then be $(R/\tau_2)Y = \beta XY$, where $\beta = R_0/(N\tau_2)$.

If, as is conventional^{1,4}, we make the further assumption (iii) that β is a constant, we are effectively assuming that R_0 is proportional to the population density N . An immediate consequence of this assumption is that a control policy, whether by vaccination or culling, will eliminate the disease if and only if the susceptible population is kept at or below the threshold population density K_T defined by $R_0 = 1$, that is, $K_T = 1/(\beta\tau_2)$ (refs 1, 2, 5). However, there are few diseases for which the assumption is quantitatively plausible. In many cases, and especially among territorial animals such as foxes, it seems reasonable to assume that R_0 will rise more slowly than linearly with population density. (R_0 would rise linearly with population density if the area over which an animal ranges were independent of density, while on a strict territorial model, R_0 could be independent of N . It seems reasonable to conjecture that the truth lies somewhere between: that is, that R_0 rises with N , but more slowly than linearly.) This has favourable implications for a vaccination control strategy, because the criterion for elimination of the disease is that the proportion unvaccinated should be less than $1/R_0$, and this can be met in habitats of greater carrying capacity if R_0 rises more slowly with density.

When we consider culling, the basic threshold result still holds: we need to achieve $R_0 \leq 1$, and if we make the conventional assumption of a βXY infection term, with β constant, the conclusion is that we must reduce the population density to below the threshold carrying capacity K_T . Our analysis in terms of model components here has an advantage of a rather negative kind. By revealing that control depends on a single easily understood ecological parameter, R_0 , our analysis makes it much easier to appreciate that our model on its own is inadequate for a consideration of culling: we cannot expect a population held down to density K_T by culling to behave like a natural population at the same density. Indeed, how it does behave will depend on the culling strategy. There is evidence that a strategy which disrupts the social pattern can be counterproductive (Ross, reported in ref. 3), increasing R_0 even as it reduces the population density. On the other hand, a relatively modest culling strategy might succeed if aimed selectively at the types of individual who make most contacts.

Second, I shall discuss some aspects which depend principally on the population growth and generation gap terms. In endemic equilibrium the net growth rate ρX must be equal to the infection rate [βXY in equation (1)], which in turn must equal the turnover rates of the incubating and infectious classes, I/τ_1 and Y/τ_2 , respectively [here τ_1 is the mean incubation period, which equals $1/\sigma$ for equation (1)]. Hence, in equilibrium the 'level of prevalence', defined as $(I + Y)/X$, will be equal to $\rho(\tau_1 + \tau_2)$. For

strongly endemic areas we may take $\rho \approx r$; also $\tau_1 + \tau_2$ is approximately (for some simple models exactly²) equal to the mean generation gap τ . Hence, the level of prevalence should be approximately equal to $\rho\tau$, and thus vary from zero in near-threshold habitats up to $r\tau$ in areas where the disease holds the population down well below the carrying capacity. This simple argument does not depend on the detailed form of any of our three epidemic model components (it is sensitive, though, to spatial heterogeneity⁶).

The period of any oscillations about equilibrium also seems to be insensitive to the detailed form of the population growth and generation gap terms, being approximately equal to $2\pi\sqrt{(\tau/\rho)}$ for models with widely varying details². (It does, however, depend on the relationship between R_0 and the population density: if this is less than linear, as suggested above, the period would be somewhat longer.) On the other hand, the stability of oscillations is found to be very sensitive to the form of both terms: a net population growth rate which falls steadily as the population density rises (for example, linearly¹), and a generation gap with considerable variation [for example, a sum of two exponentials, as in equation (1)], are both conducive to stability. Other factors will also affect stability, for instance, seasonal and stochastic variation (both probably destabilizing^{7,8}) and local rather than homogeneous mixing (probably stabilizing⁶).

Although ostensibly restricted to host-regulating diseases, the analysis presented here, especially the discussion of control, will have wider applications to models incorporating similar components. For example, in endemic equilibrium we must still have $R = 1$, and hence $X = N/R_0$ if we still assume homogeneous mixing (assumption (ii) in the discussion of the infection term). Thus, in a non-fatal disease such as measles, where N and hence R_0 will be unchanged, the equilibrium proportion of susceptibles will also be unchanged. Our approach makes it clear that this result is not related to the dependence of R_0 on N , but depends only on the success probability of contacts remaining the same before and after vaccination.

Received 5 November 1983; accepted 8 May 1984.

1. Anderson, R. M., Jackson, H. C., May, R. M. & Smith, A. M. *Nature* **289**, 765–771 (1981).
2. Mollison, D. in *Mathematical Aspects of Rabies Epizootics* (ed. Bacon, P. J.) (Academic, London, in the press).
3. Bacon, P. J. & Macdonald, D. W. *New Scientist* **87**, 640–645 (1980).
4. Bailey, N. T. J. *The Mathematical Theory of Infectious Diseases* 2nd edn (Macmillan, New York, 1975).
5. Dietz, K. in *Epidemiology* (eds Ludwig, D. & Cooke, K. L.) 104–121 (SIAM, Philadelphia, 1974).
6. Mollison, D. & Kuulasmaa, K. in *Mathematical Aspects of Rabies Epizootics* (ed. Bacon, P. J.) (Academic, London, in the press).
7. Bartlett, M. S. *Stochastic Population Models* (Methuen, London, 1960).
8. Stirzaker, D. R. *J. Inst. Math. Jussieu* **15**, 135–160 (1975).

A novel cold-tolerant insect found in a Himalayan glacier

Shiro Kohshima

Department of Zoology, Kyoto University,
Kitashirakawa-Oiwakecho, Sakyo-ku, Kyoto 606, Japan

Here we report the discovery of a new species of cold-tolerant midge (Chironomidae, *Diamesa* Meigen sp.) in a high-altitude glacier of the Nepal Himalayas. The adult insect, characterized by reduced wings and antennae (Fig. 1a), is unable to fly, and is found walking on the surface of the glacier and in small cavities beneath it. The larvae grow in melt-water drainage channels under the ice and feed on blue-green algae and bacteria. The insect is the first to be found which spends its entire life cycle in the snow and ice of a glacier—the coldest insect habitat ever recorded^{1,2}. The insect was active at temperatures as low as -16°C , well below those at which activity has been seen in insects living in other cold habitats, including Antarctic ones. The study also reveals a previously unsuspected ecosystem based on the algae and bacteria growing on glacial ice.

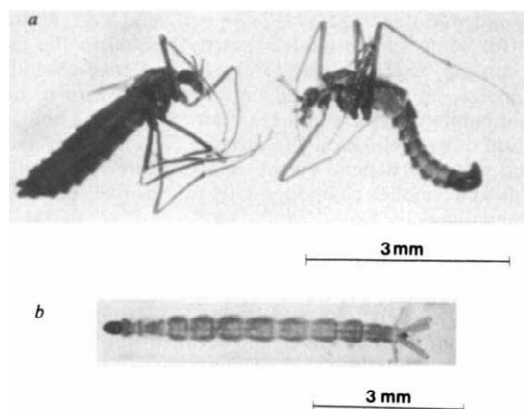


Fig. 1 a, Adults of *Diamesa* sp. Left, female; right, male. b, Larva of *Diamesa* sp.



Fig. 3 Section of an under-snow melt-water drainage channel of the ablation area of the glacier (along the dotted line).

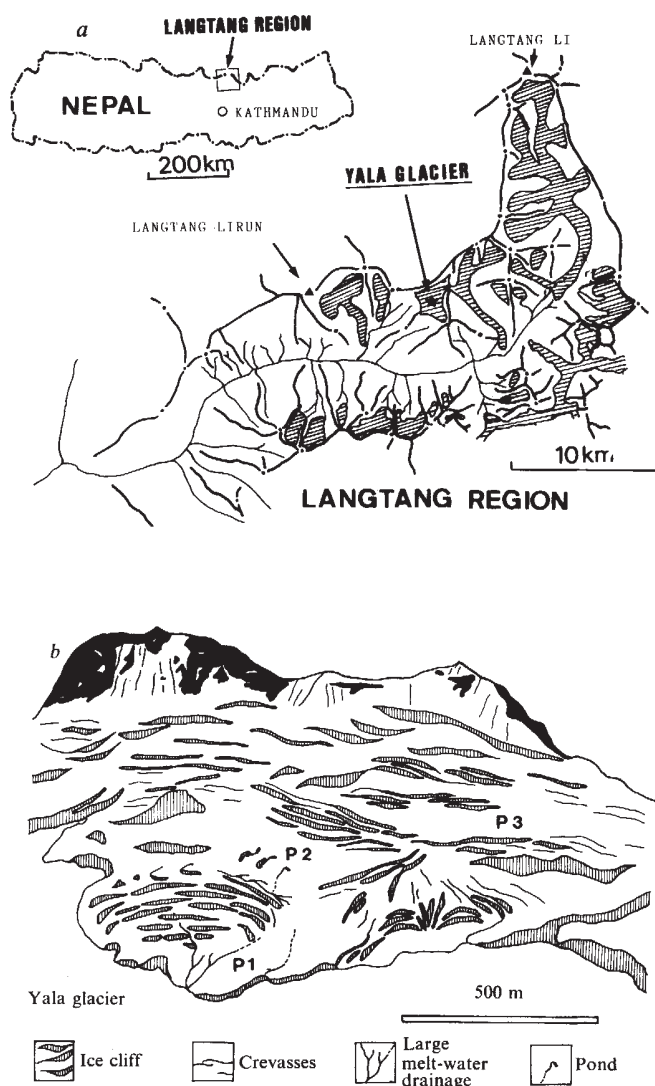


Fig. 2 a, Map of the Langtang region, showing the location of the Yala glacier. b, Outline map of the Yala glacier (5,100–5,600 m altitude).

Research was done at the Yala glacier (5,100–5,600 m altitude) in the Langtang region of Nepal between 14 September and 24 October 1982, from the end of the monsoon season to the beginning of winter. The Yala glacier is a clean mountain glacier, without debris, and has many flat terraced plateaus divided by ice cliffs and crevasses (Fig. 2). The study was concentrated on three of these plateaus, termed P1, P2 and P3 (5,130 m, 5,200 m and 5,400 m altitude, respectively), at a season when each was covered with snow (0.1–1.0 m on P1, 1.0–2.0 m on P2 and >3 m on P3) and had many tunnel-like melt-water drainage channels running along the boundary of the glacial ice and the snow cover (Fig. 3).

The only insect found on the Yala glacier was the previously unknown midge *Diamesa* Meigen sp. The main habitats of the insect are melt-water drainage channels and small cavities in the snow and ice. In the daytime, many adults were seen walking on the surface of the snow when the Sun was shining, but they would quickly disappear when it became overcast, going down through the snow to the surface of the glacier ice (ablation surface).

Larvae (Fig. 1b) were found living in the melt-water drainage channels running along the ablation surface of P1 and P2. The ablation surface in this part of the glacier contained many characteristic small pits, each with small mud-like granules (0.1–2.0 mm in diameter) at their bottoms. At the bases of the melt-water drainages on P1 and P2, the pits were often as much as 10 cm wide and 30 cm deep and housed large numbers of larvae and granules. Microscopic observation revealed that the main components of the granules were living micro-plants such as aquatic blue-green algae (*Phormidium* sp.) and bacteria³. Dissection of the intestines of many larvae revealed that they were feeding on these granules.

Almost all the adults observed on the snow surface were females: 1,008 individuals out of a total of 1,012 (99.6%). Most adult males, and all copulating pairs, were found on the ablation surface beneath the snow. It thus seems that males spend virtually their entire lives beneath the snow surface. In contrast, adult females were commonly found on the surface and appeared to be able to orient by means of a Sun compass. They normally walked in straight lines, the direction of which could be altered by changing the apparent position of the Sun with a hand mirror. A similar phenomenon is seen in a Japanese wingless stone-fly, *Eocapnia nivalis*⁴.

Diamesa lives at remarkably low temperatures. Figure 4 shows the diel change in air and snow temperatures and the activity of adult insects recorded on P1 from 14 to 15 October 1982.

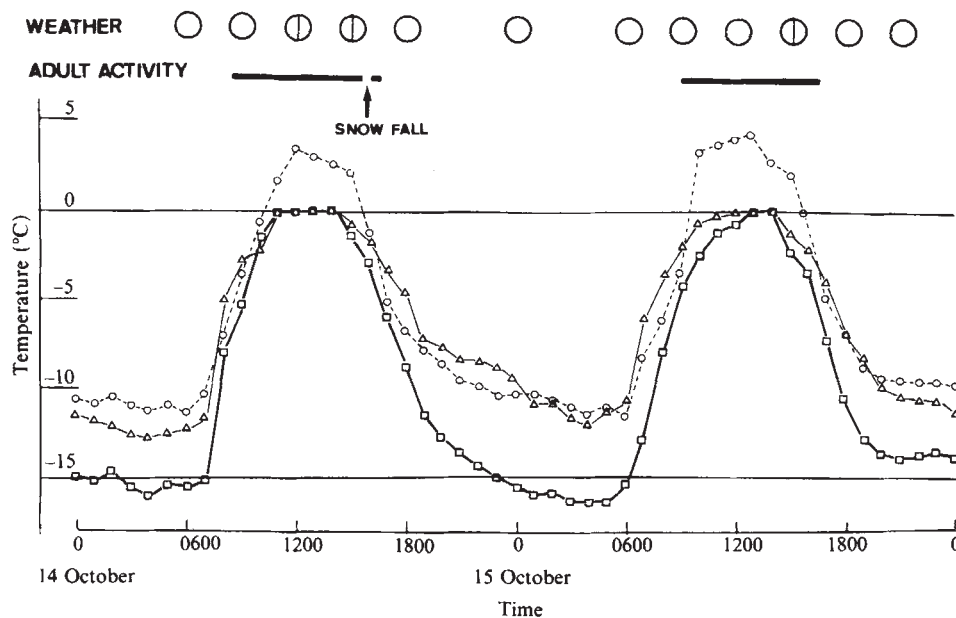


Fig. 4 Diel change in air and snow temperatures and activity of adult insects recorded at P1 on 14–15 October 1982. ○—○, Air temperature (80 cm above the snow surface); □—□, temperature of the snow surface; △—△, temperature of the ablation surface under 10 cm snow cover (recorded by K. Takahara). Weather symbols: ○, very fine (less than 20% of the sky covered with clouds); ⊕, fine (20–70% of the sky covered with clouds).

Activity was observed during the day time when the snow surface temperature ranged from 0.0 °C to −7.2 °C and the temperature of the ablation surface (10 cm under the snow), where all the adults were at night, ranged from 0.0 °C to −12.8 °C.

Adults of the species have a unique and remarkable ability to move and function at very low temperatures. Adults were captured with an insect aspirator and placed in plastic cups which were transported to P3 where they were kept on the snow in a small snow house in order to observe their activity free from any warming from the Sun. The lowest temperature at which adult activity was recorded was −16 °C (at 0600 h, before sunrise on 29 September). The insects were at first inactive in the darkness, but started to move and walk slowly when illuminated by an electric torch. The same response was often recorded at night (1900–2200 h) at temperatures ranging from −10 °C to −14 °C. In view of the fact that temperatures in this range normally cause cold stupor, even in Antarctic insects^{5,6}, this finding is amazing. In contrast to their strong cold tolerance, adults of the species were very sensitive to a temperature as high as that of the human hand. When placed on a human palm, they became hyperactive for a few seconds, but were paralysed within about 20 s; when returned to the snow surface, they recovered and began to walk again.

Many insects adapted to cold environments have been reported from the Arctic and Antarctic regions^{4,5} but there has been no record of any species which exhibits normal adult activity and larval development in such a cold environment as *Diamesa* does on the Yala glacier. For example, *Belgica antarctica*, an apterous chironomid living in the Antarctic Peninsula, is the southernmost free-living holometabolous insect, but adult activity and larval growth is seen on the ground only after the disappearance of snow in the austral summer when the air temperature ranges from −3 °C to +10 °C.

Head capsule widths of the *Diamesa* sp. larvae collected from 22 to 28 September show a uni-modal distribution (mean = 0.36 mm, range 0.30–0.42 mm, $n = 88$), indicating that the insect has at least one generation per year. Emergence of pupae and adults was observed during the entire study period (18 September–24 October), but the number of larvae and pupae decreased with the advancing season; no larvae and only a few pupae were found with many pupal exuviae on 19 October. This indicates that almost all larvae reached the last instar at the end

of September and completed metamorphosis by the end of October. It seems unlikely that the final instar larvae overwinter in the drainage channels as do other chironomid species.

The lifespan of adult females of this species is exceptionally long for a chironomid (at least 1 month). 100 adults were kept without food in plastic cups on the snow on P1 and in the snow houses on P3 for 35 days from 19 September to 24 October. Out of 50 females, 46 remained alive but all 50 males died. Dissection of the captive females and the field specimens collected at the end of October showed that they still contained immature eggs and well developed fat bodies even though winter was just beginning. This strongly suggests that females overwinter in the adult stage after migration over the snow surface. That females could survive without food and the fact that field-collected adults had empty intestines indicates that this species does not feed in the adult stage. This is an exceptional life cycle for a chironomid, most of which have a very short adult period and overwinter in their larval stage. However, for an insect living on moving glacial ice on which the distribution of melt-water drainages changes annually, it may be a better strategy to overwinter in the adult state and oviposit the next spring in a well established drainage channel, because the pattern of the spring water flow cannot be predicted at the beginning of the previous winter.

This work was supported by a Grant-in-aid for Overseas Field Research no. 57041023 awarded to Dr K. Higuchi of Nagoya University and also aided in part by a Grant-in-Aid for Special Project Research no. 58121003 from the Ministry of Education, Science and Culture of Japan. I thank Dr O. A. Saether for identifying the insect specimens and Dr M. Akiyama for identifying the algal specimens; the members of the research team (Glaciological Expedition to Nepal 1982) for their support and warm encouragement; and Dr T. Hidaka for helpful advice and for reading the manuscript.

Received 4 December 1983; accepted 29 May 1984.

- Swan, L. W. *Scient. Am.* 205, 68–78 (1961).
- Mani, M. S. *Introduction to High Altitude Entomology* (Methuen, London, 1962).
- Kohshima, S. in *Glacial Studies in Langtang Valley* (ed. Higuchi, K.) 91–97 (Data Center for Glacier Research, Japanese Society of Snow and Ice, Nagoya University, 1984).
- Kohshima, S. & Hidaka, T. *Biol. Inland Wat.* 2, 19–23 (1981).
- Sugg, P., Edwards, J. S. & Baust, J. *Ecol. Ent.* 8, 105–113 (1983).
- Oliver, D. R. A. *Rev. Ent.* 16, 211–230 (1971).

Continuous *in vitro* generation of multipotential stem cell clones from src-infected cultures

Elaine Spooncer, David Boettiger* & T. Michael Dexter

Department of Experimental Haematology, Paterson Laboratories, Christie Hospital & Holt Radium Institute, Withington, Manchester M20 9BX, UK

A molecular recombinant of Rous sarcoma virus and murine amphotropic leukaemia virus, src(MoMuLV), where the avian src oncogene has been placed under the influence of a murine virus promoter sequence, has been reported¹. Infection of long-term marrow cultures² with this virus led to a dramatic change in the relative numbers of stem cells, granulocyte-macrophage progenitor cells and mature cells found in normal haematopoietic cell development³. However, although the balance between self-renewal, differentiation and development was disturbed, injection of the cultured cells into irradiated syngeneic recipients did not lead to the development of leukaemia. Thus, although the control had been 'loosened', the host regulatory mechanisms were sufficient to impose a restraint on unlimited growth of the cells. We now show that the stem cells from the src-infected cultures show a remarkably increased capacity for self-renewal *in vitro* in situations which are inimical to the maintenance of self-renewal in normal uninfected stem cells and that self-renewal/differentiation can be modified by the culture conditions.

We have reported previously that stem cells (CFU-S) from src (MoMuLV)-infected long-term marrow cultures, unlike normal CFU-S⁴⁻⁶, possess an ability to undergo a sustained serial transfer *in vivo*³. This indicated that the self-renewal probability of the stem cells had been altered in some way. Self-renewal of stem cells from the src (MoMuLV)-infected and from control long-term cultures has now been tested by their ability to transplant serially *in vitro* on the microenvironment provided by the adherent layer of long-term marrow cultures (Table 1). In agreement with previous studies⁷⁻⁹, CFU-S from control long-term cultures possess a low self-renewal capacity and rapidly decline when serially transferred *in vitro*. CFU-S from the src (MoMuLV)-infected cultures, on the other hand, can readily be maintained over several *in vitro* transfers. The stem cells (CFU-S) produced in this serial transfer system are apparently 'normal' in that they produce spleen colonies containing all the different cells of the myeloid lineages; they can protect mice from the effects of potentially lethal radiation; the reconstituted mice do not develop a leukaemia, although a slight anaemia (haematocrit ~37) is quite common. In the serially transferred cultures, granulocyte-macrophage progenitor cells (GM-CFC) are also being produced but there has been an alteration in the ratio of CFU-S:GM-CFC from about 1:3 in the original src (MoMuLV)-infected cultures to between 1:22 and 1:53 in the serially transferred cultures. This supports our previous suggestion³ that the developmental 'block' seen in the original infected cultures is due to a progressive transformation of the adherent (stromal) cells—an effect which is abrogated when the haematopoietic cells are transferred to a new environment. However, the increased capacity for self-renewal of stem cells upon serial transfer *in vitro* suggests further that the CFU-S from the src (MoMuLV)-infected cultures have undergone some intrinsic change in self-renewal ability. In these serially transferred cultures, there is production of infectious virus and expression of high levels of src kinase activity³. However, whether the increased self-renewal of stem cells is due directly to infection of these cells with the src (MoMuLV) and expression of kinase activity has been difficult to determine because of the relatively low numbers of stem cells and the presence of stromal cells, which may be the true target cells for infection, production and expression of virus.

Table 1 Serial transfer of src (MoMuLV)-infected and control cultured stem cells onto irradiated marrow stroma

	CFU-S transferred	CFU-S recovered	GM-CFC recovered	Ratio CFU-S: GM-CFC
Control				
Original	(180)	—	(3,400)	(1:19)
Passage 1	1,500	150	2,800	1:19
2	150	<20	<100	1:5
3	<20	0	—	—
src (MoMuLV)				
Original	(2,500)	—	(8,500)	(1:3)
Passage 1	1,750	1,287	47,619	1:37
2	644	1,480	63,640	1:43
3	740	368	8,096	1:22
4	184	190	10,830	1:27
5	95	156	8,268	1:53

Long-term bone marrow cultures from adult B6D2F₁ mice were established and infected with $1-2 \times 10^4$ focus-forming units of src (MoMuLV), 5-8 weeks after the start of the cultures, as previously described³. The cultures were maintained on a weekly demi-depopulation feeding regime in Fischer's medium (Gibco) containing 20% (v/v) horse serum (Flow) and 10^{-6} M (final concentration) of hydrocortisone sodium succinate and maintained at 33 °C (ref. 2). After 24 weeks of culture, the supernatant cells from control and from src (MoMuLV)-infected cultures were removed, counted and assayed for CFU-S and GM-CFC as previously described³. These data are shown as the numbers of stem cells and GM-CFC present in the original culture shown. To assay for CFU-S, between 2×10^3 and 10^5 cells were injected intravenously into syngeneic mice that had received 15 Gy of radiation from a ⁶⁰Co source. Spleen colonies were counted 9-11 days later. GM-CFC were assayed by plating 10^3-10^5 cells per ml in soft-agar gels containing WEHI-3 cell conditioned medium as a source of CSF (refs 1, 9). At the time of assay of the original cultures for CFU-S and GM-CFC, aliquots of the supernatant cells were inoculated into culture flasks containing a previously established normal marrow cell adherent layer. To do this, marrow cells were inoculated into culture flasks and allowed to form an adherent layer over 3 weeks of culture⁷. At this time they were subjected to 15 Gy γ -radiation using a ¹³⁷Cs source. One day later, the growth medium (containing dead and dying haematopoietic cells) was removed and replaced with fresh medium. In these cultures, endogenous haematopoiesis is totally ablated by the radiation. These cultures were then used within 1 week as recipient adherent layers for the cultured control or src (MoMuLV)-infected cells. Thus, in the first passage of the control cells, the inoculated cells contained 1,500 CFU-S and in the first passage of the src (MoMuLV)-infected cells, 1,750 CFU-S were added to the irradiated adherent cell layer. The cultures were then maintained for 2 weeks on a weekly demi-depopulation feeding regime. After 3 weeks, the supernatant cells were collected and assayed for CFU-S and GM-CFC (total recovered per flask) and the results expressed as a ratio. The cells were also transferred onto fresh irradiated adherent cell layers, established precisely as described above, and the procedure repeated at 3-weekly intervals (passages 1-5). Because the cultures had been fed for 2 weeks before the assay and transfer point, the data shown in the table are an underestimate of the total cumulative CFU-S and GM-CFC produced in the cultures throughout each successive transfer. On several occasions, cells from the primary src (MoMuLV)-infected cultures and from the serially transferred cultures were tested for their leukaemic potential. Recipient mice were subjected to potentially lethal radiation and injected with 5×10^3 to 5×10^6 cells. These animals showed complete haematopoietic regeneration. There were a total of 56 mice in this study; none has developed a leukaemia, and the longest-lived survivor is now 14 months past reconstitution.

To test this, we investigated the capacity of the src (MoMuLV)-infected cells to undergo extended self-renewal in the absence of the stromal cells normally required for haematopoiesis *in vitro*^{2,8,10}. We have serially transferred colonies developing in soft-gel media using both the GM-CFC assay (which is 'biased' towards detection of granulocyte/macrophage progenitors) and the CFC-mix (mixed colony-forming cells) assay which facilitates the clonogenic growth of multipotential cells¹¹⁻¹³—at least some of which are equivalent to the *in vivo* assayed CFU-S (ref. 11). Because of the importance of

* Department of Microbiology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA.

Table 2 Serial transfer of control and src (MoMuLV)-infected cells in a 'GM-CFC' assay

Transfer number	No. of colonies transferred	Stimulus used	No. of colonies recovered	Cumulative total colony increase
Control		0	0	0
Original	(10 ⁵ cells)	HCGF	164	—
		CSF-1	150	—
	164	HCGF-1	1	0.01
	(from HCGF)			
	150	CSF-1	0	—
	(from CSF-1)			
src (MoMuLV)				
Original	(4 × 10 ³ cells)	HCGF	74	—
		CSF-1	1	—
1	17	HCGF	800	47
	(from HCGF)			
	3	CSF-1	0	0
	(from CSF1)			
2	8	HCGF	214	1.3 × 10 ³
3	0.5	HCGF	595	1.4 × 10 ⁶
4	6	HCGF	256	6.1 × 10 ⁷
5	2.6	HCGF	250	6.0 × 10 ⁹
6	2.5	HCGF	156	3.7 × 10 ¹¹
7	1.6	HCGF	300	7.1 × 10 ¹³
8	3.0	HCGF	209	5.0 × 10 ¹⁵
9	2.0	HCGF	394	9.4 × 10 ¹⁷
	50.0	0	0	—

Long-term cultures infected with src (MoMuLV) or control cultures were established as described in Table 1 legend. 20 weeks after infection, the cultured cells were plated in 1-ml agar gels consisting of Fischer's medium supplemented with 20% horse serum, 0.3% final concentration of agar, and either no stimulus, 15% (v:v) WEHI-3 cell conditioned medium as a source of HCGF (ref. 12) or 15% (v:v) of mouse L-cell conditioned medium as a source of CSF-1 (ref. 16). The agar gels were incubated for 7 days at 37 °C in fully humidified conditions containing air plus 5% CO₂, and the numbers of discrete colonies (>50 cells) were determined 7 days later. Each assay was performed in triplicate and the results are the mean of three plates. Standard errors fell within the range ±20%. After counting the plates, the whole agar gel from one plate (1 ml) containing colonies was resuspended in 5 ml of growth medium and the colonies were dispersed by repeated flushing with a 1-ml pipette. The cell suspension was then further diluted as appropriate and fractions of this were plated as 1-ml aliquots in the required stimulatory material (either HCGF or CSF-1). This was repeated at weekly intervals. The results are expressed as a fraction of the number of colonies transferred based upon colony counts at day 7 (ref. 1). However, on three different occasions (transfers 1, 5 and 9) we picked out individual colonies and replated these cells separately. Of a combined total of 112 colonies assayed in this way, 70% contained further colony forming cells and the mean increase gave results similar to that seen by replating fractions of the whole gel. That the cells are responding to HCGF, rather than to some other stimulatory material, was shown on two separate occasions when the cells were replated in a highly purified HCGF preparation¹⁴. Equivalent colony counts and replating efficiency were obtained in each case (data not shown). On repeated occasions, the cells developing in these colonies were tested for their ability to survive and proliferate in the absence of HCGF. No survival or growth was seen when the cells were plated between 10³ and 2 × 10⁶ per ml in the absence of HCGF.

different growth factors in progenitor and stem cell self-renewal^{10,14}, the assays were performed either using WEHI cell conditioned medium (as a source of the multi-lineage and stem cell stimulating factor which we call haematopoietic cell growth factor (HCGF) (ref. 12, see also refs 13, 15) or using L-cell conditioned medium (as a source of the macrophage-specific stimulus called CSF-1; ref. 16).

Table 2 shows that the control long-term cultured cells respond equally well to either HCGF or CSF-1 in terms of colony formation in the 'GM-CFC' assay system but that few or no colony-forming cells are present after only 7 days of growth in soft-agar gels. These data agree with those previously published^{3,14}. Cells from src cultures, on the other hand, still

Table 3 Serial transfer of control and src (MoMuLV)-infected cells in a 'CFC-mix' assay

Transfer number	No. of colonies transferred	No. of colonies recovered	Cumulative total of colony increase
Control			
Original	(10 ⁵ cells)	94	—
1	4	0	0
src (MoMuLV)			
Original	(2 × 10 ³ cells)	50	—
(18 weeks)			
1	5	36	7.2
2	3.6	39	78
3	3.9	6.6	132
4	1.32	0	132
src (MoMuLV)			
Original	(1.2 × 10 ³ cells)	30	—
(20 weeks)			
1	3	24	8
2	2.4	12	40
3	1.2	0	40

The cells from src (MoMuLV)-infected or control cultures were plated as for Table 2 legend except that the culture medium consisted of Iscove's modified Dulbecco's (Gibco), 20% (v:v) fetal calf serum, 1% (w/v) bovine serum albumin, 10% WEHI-CM as a source of HCGF (multi-CSF; refs 12–14), 1 U ml⁻¹ erythropoietin (Hyclone) and 0.3% (final concentration) agar (method based on ref. 11). The plates were incubated in fully humidified conditions at 37 °C in an atmosphere of 5% O₂, 5% CO₂ in nitrogen. Colony counts were made at day 7 and again at days 12–14. Individual colonies were picked out from the agar at days 12–14, cytopsin preparations made, and stained with benzidine plus May-Grunwald/Giemsa before microscopic analysis of the cell types present. In the colonies picked out from control cultures, up to 10% contained cells of more than one lineage (erythroid cells with either granulocytes or megakaryocytes). Of colonies picked out from src (MoMuLV) cultures (either primary or transferred), 20–70% of the colonies contained multiple cell lineages. The remaining colonies contained granulocytes and/or macrophages. Colony transfers were performed at 7-day intervals exactly as described in the legend to Table 2 except that the different growth medium was used. Experiments were performed on the src (MoMuLV)-infected cells at 18 weeks and 20 weeks after infection. Note that a similar decline was seen in either instance.

require growth factor for colony formation, but note the disparity in the response to HCGF versus CSF-1. With CSF-1, few colonies are formed and these are not replatable. In HCGF, however, colony-forming cells are continually being generated even after the ninth successive transfer. At this time, the colony forming cells still absolutely require the presence of HCGF for expression of their growth potential. Morphological examination of the colonies was performed at each successive transfer. All colonies examined contained recognizable mature neutrophils—but all contained a high proportion (10–90%) of primitive undifferentiated cells and frequent mitosis.

The presence of large numbers of undifferentiated cells in these colonies indicated that the assay system may not be allowing full expression of the differentiation potential of the cultured cells. Therefore, we performed a similar experiment where the cells were cloned and replated in an assay system promoting the development of multipotential cells¹¹. The results (Table 3) are of interest from several points of view. First, in the primary plates of the cells taken directly from the src-infected, long-term cultures, the same number (if not more) of colonies formed in the CFC-mix, multipotential cell assay, compared with the GM-CFC assay system. Moreover, of colonies developing in the CFC-mix system, 20–70% contained maturing cells of more than one myeloid lineage (erythroid cells with granulocytes and/or megakaryocytes). Second, the replating efficiency of these cells (for their ability to produce both single-lineage and mixed-

Table 4 Multipotential stem cell production from src (MoMuLV)-infected cells in a 'GM-CFC' assay

Transfer number	No. of colonies transferred	No. of colonies recovered	% of mixed myeloid colonies
1	0.7	45	30
3	0.25	140	ND
4	0.23	20	50
6	0.32	80	33
10	4	287	30

Long-term marrow cultured cells taken from src (MoMuLV)-infected cultures, 27 weeks after infection, were serially passaged at 7-day intervals in a GM-CFC assay system exactly as described in the legend to Table 2. These cultures also gave a cumulative colony increase similar to that shown in Table 2. However, at selected passages (transfer numbers, that is, the number of weeks since the primary plating) whole gel cell suspensions were made and assayed for the presence of multipotential cells using the modified growth conditions as described in the legend to Table 3. The number of colonies was counted at days 12–14, and individual colonies were picked out, stained and classified as to the lineages present. Mixed colonies contained benzidine-positive nucleated and enucleate erythroid cells plus cells of one more lineage (granulocytes, macrophages or megakaryocytes).

lineage colonies), although obviously greater than the corresponding control cells, does not show the prolonged maintenance demonstrated in the GM-CFC assay system, that is, the probability has been somewhat 'shifted' in favour of differentiation rather than self-renewal. In support of this, morphological analysis of the developing colonies in the CFC-mix system has demonstrated a spectrum of maturing cells not seen in the GM-CFC system. Although in the majority of these colonies, also, undifferentiated cells are still present, these are generally in a lower proportion (<10%) than that seen in colonies developing in the GM-CFC assay. Thus, the larger number of blast cells seen in the GM-CFC assay system probably represents a developmental block occurring in these culture conditions. Consequently, the real developmental potential of the cells is seen only when they are plated in conditions which facilitate multi-lineage development.

The differences seen in the ability of the cells to undergo repeated serial recloning in the two assay systems is not simply a selection of a different cell population or the emergence of the unusual 'factor-dependent' (IL-3 or HCGF dependent) cell lines previously described^{17–19}. Rather, it appears to be a reflection of the capacity of the different culture conditions to maintain self-renewal of multipotential cells (Table 4). Colonies developing in the GM-CFC assay system were tested at different transfer numbers for their ability to produce mixed myeloid colonies in the CFC-mix system. The results clearly show that within the GM-CFC assay system, large numbers of multipotential cells are produced and maintained at each serial transfer. Therefore, we conclude that the self-renewal observed in the GM-CFC assay is a reflection of the different culture conditions.

It is clear from these results that infection of long-term cultures with src (MoMuLV) has facilitated the growth of stem cells with an unusual capacity for self-renewal when serially transferred on marrow stroma. We also show that the increased self-renewal ability of multipotential cells in the CFC-mix assay from src (MoMuLV) cultures is maintained in the absence of stroma, provided that the cells are supplied with a multi-lineage stimulating factor, HCGF. Interestingly, self-renewal in this case can be modified further by the culture conditions. The efficiency of self-renewal seems to be inversely related to the differentiation pressure imposed by the culture system but the role of pp60^{src} (ref. 20; the product encoded by the src oncogene) in the maintenance of multipotential cells in the absence of stroma remains an enigma. Previous work has shown³ that long-term

cultures infected with src (MoMuLV) both produce infectious virus and express high levels of src kinase activity. These results have been confirmed by J. Wyke and A. Stoker (personal communication). Also, the expression of the src oncogene in the primary infected long-term cultures has had a role (directly or indirectly) in the events observed, as similar effects have not been seen with stem cells from long-term cultures infected with other helper or rapidly transforming RNA viruses carrying other oncogenes^{21–23}. However, we have recently taken individual colonies developing in the soft-gel systems described here and amplified the cells in liquid culture in the presence of HCGF. These cell lines contain multipotential cells which will form mixed myeloid colonies when plated in soft-agar and they can also establish haematopoiesis and produce CFU-S when inoculated onto marrow stromal cells (E.S. and T.M.D., manuscript in preparation). It seems likely, therefore, that some if not most of the colonies developing in our serial transfer *in vitro* systems are derived from the most primitive of haematopoietic stem cells, the CFU-S, that is, src (MoMuLV) infection of long-term marrow cultures has provided an opportunity for the culture of non-leukaemic, multipotential stem cells. Surprisingly, however, the production of infectious virus or expression of high levels of src kinase activity by these cells has not been detected so far (J. Wyke and A. Stoker, personal communication). One possibility is that src (MoMuLV) infection of the original marrow cultures 'selected' for stem cells with a high self-renewal ability, but which were themselves refractory to infection with the virus. In this case, however, it is difficult to envisage how the stability of the phenotype is maintained through several *in vitro* reclonings. Alternative possibilities include a 'hit-and-run' event or the integration of provirus adjacent to some important regulatory gene. If true, these must be highly frequent occurrences to account for the ease with which such 'transformation' events are seen in the cultures. It is also possible that culture of cells in the presence of HCGF in some way modulates the production and expression of virus. These possibilities are being explored.

Certainly, the problem of self-renewal lies at the heart of normal haematopoietic cell regeneration as well as leukaemogenesis. Thus far, the problem has been difficult to approach because of the lack of suitable *in vitro* systems in which events regulating growth and differentiation can be examined in a non-malignant 'normal' self-renewing system with a capacity for multiple lines of differentiation. The present results suggest that these cultures may be very useful in this respect.

This work was supported by the Cancer Research Campaign (UK).

Received 14 March; accepted 15 May 1984.

- Anderson, S. M. & Scolnick, E. M. *J. Virol.* **46**, 594–605 (1983).
- Dexter, T. M., Allen, T. D. & Lajtha, L. G. *J. cell. Physiol.* **91**, 335–344 (1977).
- Boettiger, D., Anderson, S. & Dexter, T. M. *Cell* **36**, 763–773 (1984).
- Schofield, R. & Lajtha, L. G. *Blood Cells* **9**, 467–473 (1983).
- Harrison, D. E. & Astle, C. M. *J. exp. Med.* **156**, 1767–1779 (1982).
- Siminovich, L., Till, J. E. & McCulloch, E. A. *J. cell. comp. Physiol.* **64**, 23–31 (1964).
- Mauch, P., Greenberger, J. S., Botnick, L., Hannon, E. & Hellman, S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2927–2930 (1980).
- Dexter, T. M. *et al.* in *Aplastic Anemia* (Liss, New York, in the press).
- Schofield, R. & Dexter, T. M. *Leuk. Res.* (in the press).
- Dexter, T. M. & Allen, T. D. *J. Path.* **141**, 415–433 (1983).
- Metcalfe, D., Johnson, G. R. & Mandel, T. E. *J. cell. Physiol.* **98**, 401–412 (1979).
- Humphries, R. K., Eaves, A. C. & Eaves, C. J. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3629–3633 (1981).
- Iscove, N. N., Roitsch, C. A., Williams, N. & Guilbert, L. J. *J. cell. Physiol. Suppl.* **1**, 65–78 (1982).
- Bazill, G. W., Haynes, M., Garland, J. & Dexter, T. M. *Biochem. J.* **210**, 747–759 (1983).
- Burgess, A. W., Metcalfe, D., Russell, S. H. M. & Nicola, N. A. *Biochem. J.* **185**, 301–314 (1980).
- Stanley, E. R. & Guilbert, L. J. *J. immun. Meth.* **42**, 253–284 (1981).
- Dexter, T. M., Garland, J., Scott, D., Scolnick, E. & Metcalfe, D. *J. exp. Med.* **152**, 1036–1047 (1980).
- Greenberger, J. S. *J. supramolec. Struct.* **13**, 501–511 (1980).
- Schrader, J. W., Lewis, S. J., Clark-Lewis, I. & Culvenor, J. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 323–327 (1981).
- Brugge, E. & Erikson, R. L. *Nature* **269**, 346–348 (1977).
- Dexter, T. M., Scott, D. & Teich, N. M. *Cell* **12**, 355–364 (1977).
- Dexter, T. M., Allen, T. D., Testa, N. G. & Scolnick, E. J. *J. exp. Med.* **154**, 594–608 (1981).
- Greenberger, J. S. *J. natn. Cancer Inst.* **62**, 337–348 (1979).

Successful transplantation across major histocompatibility barrier of deoxyguanosine-treated embryonic thymus expressing class II antigens

A. R. Ready, E. J. Jenkinson, R. Kingston
& J. J. T. Owen

Department of Anatomy, University of Birmingham, Medical School,
Vincent Drive, Birmingham, B15 2TJ, UK

Foreign tissues grafted into healthy recipients are usually rejected by the hosts' immune system largely by means of major histocompatibility complex (MHC) products expressed on donor cells¹. During ontogeny, developing T lymphocytes acquire tolerance to self-MHC antigens and the thymus has been considered as the most likely site for the abrogation of self-reactive clones². We demonstrate here that embryonic thymus lobes, when organ cultured in the presence of deoxyguanosine, which is toxic to proliferating embryonic thymic lymphocytes but does not affect the epithelial framework, when transplanted to the kidney capsule of normal healthy histoincompatible mice, are not rejected despite their continued expression of both class I and class II donor MHC products³ but do not induce tolerance. This suggests that immunogenicity is not solely a function of MHC antigen expression but is also influenced by the type of cell upon which the antigens are expressed and, if the thymus is involved in the induction of tolerance to self-MHC products, this is a function of a component other than the epithelium, perhaps thymic dendritic cells.

Elevated blood levels of deoxyguanosine in purine nucleoside phosphorylase deficient individuals have been implicated in T-cell immunodeficiency⁴, and we have previously shown that deoxyguanosine is toxic to proliferating embryonic thymic lymphocytes, but does not affect the epithelial framework of the thymus which survives this treatment⁵. Remarkably, deoxyguanosine-treated lobes not only survive when transplanted to histoincompatible recipients but undergo considerable growth and become fully lymphoid as a result of colonization by host-derived lymphoid stem cells. Such lobes also appear to be functional *in vivo* because, when transplanted into athymic T-cell deficient nude mice, they form lymphoid grafts colonized by host derived precursors, and the thymus-dependent areas of the host peripheral lymphoid tissues (unpublished observations) are reconstituted. During a series of experiments designed to confirm the recolonization of deoxyguanosine-treated thymus *in vivo*, we observed that treated thymus grafted from CBA (H-2^k, Thy 1.2) mice to AKR (H-2^k, Thy 1.1) recipients survived and became fully lymphoid, whereas non-treated grafts were rejected after 1 week, presumably as a result of the many minor histocompatibility differences between these mouse strains (Table 1).

To determine whether deoxyguanosine treatment would also allow survival of grafts transplanted across a major histocompatibility barrier, C57BL/10 (H-2^b) mice were grafted with thymus lobes from CBA/Ca (H-2^k) mice, with or without prior deoxyguanosine treatment. Table 2 shows that, whereas untreated grafts are rejected, treated grafts survive and grow in the same way as they do in syngeneic recipients.

We have demonstrated previously that the epithelial cells of the thymus express class II MHC antigens from at least the 14th day of gestation³ and that this expression is maintained after deoxyguanosine treatment *in vitro*⁵. To examine the MHC status of thymus after growth in an allogeneic host, frozen sections were prepared from CBA/Ca deoxyguanosine-treated thymus recovered from C57BL/10 recipients 7 weeks after transplantation. Figure 1 shows that the grafts strongly express class II MHC antigens of the donor haplotype (Ia^k), thus they are present on the stromal cells of the thymic grafts without preju-

Table 1 Effect of deoxyguanosine treatment on the survival of transplanted fetal thymus lobes

Duration of graft (weeks)	No. of animals with surviving grafts	
	Treated	Untreated
1	5/5	5/5
1	5/5	0/5
3	5/5	0/5
4	4/4	0/4
5	5/5	0/5
8		

Fetal thymus lobes were isolated from CBA/Ca(H-2^k, Thy 1.2) embryos at day 14 of gestation and organ cultured for 5 days in the presence of 1.35 mM deoxyguanosine as described previously⁵. Non-treated lobes were cultured in an identical manner except for the omission of deoxyguanosine. Groups of 4–6 lobes were transplanted beneath the kidney capsule of AKR (H-2^k, Thy 1.1) recipients which were killed at weekly intervals thereafter. The presence or absence of grafts was scored using a dissecting microscope and samples were taken for histology. Successful grafts were readily identified as fleshy, white subcapsular swellings. Absence of grafts was always confirmed histologically. Grafts of embryonic thymus, whether deoxyguanosine treated or not, survived indefinitely in syngeneic recipients or nude mice (data not shown).

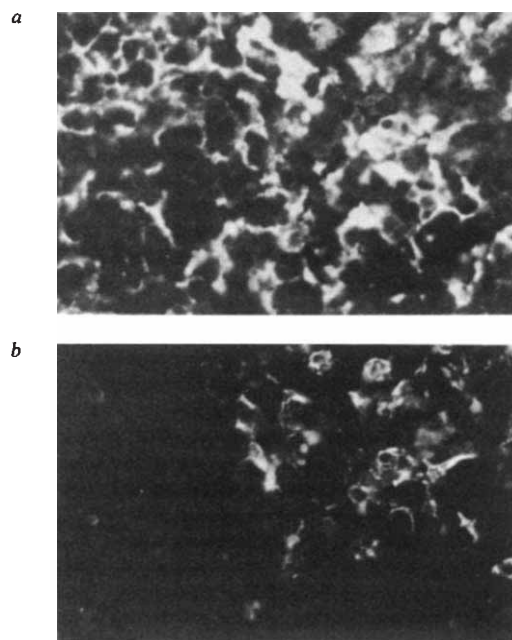


Fig. 1 Double indirect immunofluorescent labelling of a frozen section of deoxyguanosine-treated CBA thymus 7 weeks after transplantation to an immunocompetent C57BL/10 host. *a*, Fluorescein labelling showing a widespread reticular network of Ia^k comparable with that seen on the cortical epithelium of normal CBA thymus. *b*, Same field showing rhodamine labelling of A2B5-positive cells present in a more restricted cluster, a pattern also typical of the normal thymus⁶. First step reagents were monoclonal anti-Ia^k (clone 10-216, mouse IgG2, a kind gift from Dr W. Van Ewijk) and monoclonal antibody A2B5 (mouse IgM, a kind gift from Dr M. Raff). Second step reagents were rabbit anti-mouse IgG2 fluorescein (Nordic) and goat anti-mouse IgM (Fc)-rhodamine (Nordic), both at a final dilution of 1/15.

dicing their survival in an allogeneic environment. We have also detected class I (H-2K) antigen of the donor haplotype at a weaker level of staining but comparable with that found in the normal thymus⁶. Thymic grafts also contain a subpopulation of stromal cells which bind a monoclonal antibody (A2B5) directed against a GQ ganglioside characteristic of neuroendocrine cells (Fig. 1). The distribution of these cells and the pattern of staining in the grafts are similar to those found in the thymus *in situ*,

Table 2 Effect of deoxyguanosine treatment on the survival of fetal thymus grafts in syngeneic or allogeneic hosts

Host	Graft type	No. of animals with surviving grafts (%)
CBA	Treated CBA	3/5 (60)
CBA	Untreated CBA	4/5 (80)
C57	Treated CBA	7/12 (58)
C57	Untreated CBA	0/5 (0)

14-Day CBA/Ca (H-2^k) fetal thymus lobes either deoxyguanosine treated or not (as described in the legend to Table 1) were transplanted beneath the kidney capsule of allogeneic or syngeneic hosts which were killed 4 weeks later and scored for the presence of grafts. The results indicated that whilst untreated grafts are rejected by allogeneic hosts, deoxyguanosine-treated grafts grow in allogeneic mice and are recovered successfully at a frequency similar to that in syngeneic recipients.

suggesting that the allogeneic thymus transplants develop normally under the renal capsule.

The origin of the lymphoid component of the grafts was investigated by labelling with monoclonal anti-Thy 1.1 and anti-Thy 1.2 antibodies, cell suspensions derived from CBA thymus transplanted to AKR hosts. Table 3 shows that deoxyguanosine-treated thymic grafts are populated by host-derived cells with no indication of surviving donor lymphoid cells. On the other hand, analysis of untreated CBA thymus grafts to five AKR recipients, showed that at 1 week after transplantation, and just before rejection, these grafts contained $94 \pm 5\%$ donor (Thy 1.2⁺) cells. These results agree with our earlier *in vitro* studies, when it was shown that deoxyguanosine treatment eliminates indigenous lymphocytes and lymphoid stem cells, so permitting colonization by donor stem cells⁵.

Although there are several explanations for the paradox that thymus grafts, rich in MHC antigens, are not rejected by allogeneic immunocompetent hosts, two main possibilities stand out. The thymus is the primary site of T-lymphocyte production⁸ and as such might represent a special case in which expression of MHC antigens by thymic epithelial grafts renders hosts tolerant to them. However we have found (Table 4) that host lymphocytes, both within the epithelial grafts and within spleens of the recipients, can respond to the MHC antigens of the thymic graft type in mixed lymphocyte culture (MLC). This result does not exclude the possibility that powerful suppressive influences are generated *in vivo* but such influences might also be expected to operate in the case of untreated grafts—which are rejected.

The second possibility is suggested by recent studies which have shown that reduction of class II MHC antigen expressing

Table 3 Detection of cells bearing either donor or host T-cell markers in deoxyguanosine-treated thymus grafts

Graft no.	Total cell count	No. Thy 1.1 (%)	No. Thy 1.2 (%)
1	200	174 (87)	0
2	210	201 (96)	0
3	220	202 (92)	0
4	185	177 (96)	0
5	203	188 (93)	0

Grafts were carefully dissected free of the kidney capsule and cells were liberated by teasing with fine knives. The resulting cell suspensions were washed and stained with a mixture of saturating concentrations of biotin-conjugated anti Thy 1.2 (clone 303-1H Becton-Dickinson, rat IgG2b) and monoclonal anti-Thy 1.1 (NEN, mouse IgM) followed by a mixture of avidin fluorescein (Becton-Dickinson) and rhodamine-goat anti-mouse IgM (FC) (Nordic) at a final concentration of 1/50 and 1/15 respectively. Cells were scored for either rhodamine or fluorescein labelling in a Zeiss universal microscope with epi-illumination and standard filter set 9. In addition to the data shown here, examination of 24 AKR animals carrying deoxyguanosine-treated CBA thymus grafts for donor-derived Thy 1.2 expressing cells at periods from 1–5 weeks after transplantations provided no evidence for the survival of such donor lymphoid cells.

cells lowers the immunogenicity of tissue grafts⁹. Our results indicate that the survival of deoxyguanosine-treated thymic grafts is not due to simple absence of class II expression. However, immunogenicity might depend upon the cell type on which class II antigens are expressed. The thymus is known to contain two major populations of class II expressing cells—epithelial cells of the thymic stroma and dendritic cells of haematogenous origin¹⁰. We have found that deoxyguanosine treatment of embryonic thymus not only eliminates lymphocytes and their precursors, but also substantially reduces and perhaps eliminates dendritic cells and dendritic cell precursors¹¹. Thus elimination of dendritic cells may be crucial for the successful transplantation of deoxyguanosine-treated thymuses although the possibility that elimination of lymphocytes is also important cannot be excluded.

These results do show, however, that the epithelium remaining in treated grafts, despite strong Ia expression, does not seem to be immunogenic and, after colonization by host stem cells, grows into a fully lymphoid thymus. The ability of host-derived thymocytes developing within the graft to produce a response to graft type stimulators in MLC (Table 4) also argues against a role for the thymic epithelium in the induction of tolerance by the deletion of clones reactive to graft type. Furthermore, we

Table 4 MLC responses of spleen and thymus graft cells from C57BL/10 (H-2^b) mice bearing deoxyguanosine-treated CBA/Ca (H-2^k) thymus grafts beneath the kidney capsule

Animal no.	Responders	Stimulators			P-value
		Host (C57)	Graft-type (CBA)	Third party (BALB/c)	
1	Graft thymocytes	192.3 $\pm$ 50.2	1,294 $\pm$ 483.8	572.5 $\pm$ 87.4	<0.05; <0.025
	Host spleen	329.4 $\pm$ 36.5	4,245 $\pm$ 509.4	2,307 $\pm$ 216	<0.001; <0.001
2	Graft thymocytes	207.9 $\pm$ 33.9	971 $\pm$ 247.7	266.5 $\pm$ 52.6	<0.01; >0.05
	Host spleen	497.3 $\pm$ 38	5,002.4 $\pm$ 532.4	2,662.7 $\pm$ 109.3	<0.001; <0.001
3	Graft thymocytes	44.7 $\pm$ 3.1	1,805 $\pm$ 232.6	599.4 $\pm$ 51.3	<0.01; <0.01
	Host spleen	640.8 $\pm$ 90.4	10,784 $\pm$ 593.8	12,365.5 $\pm$ 911.6	<0.001; <0.001

MLC cultures were carried out in 20- μ l hanging drops in inverted Terasaki plates as described by Knight¹². Stimulators were irradiated (2,000R) spleen cells at a final concentration of 8×10^4 per well. Responder cells were prepared by gently teasing apart host spleen and thymus grafts and were used at a final concentration of 10^5 cells per well for thymus grafts and 2×10^4 cells per well for spleen. All cultures were carried out in supplemented Dulbecco's modified Eagle's medium with 10% fetal calf serum and 5% EL-4 culture supernatant as a source of T-cell growth factors¹³. Cultures were pulsed with 2 μ Ci ml⁻¹ of ³H-thymidine for 6 h on day 4 and collected for counting on an automatic cell collector. Results are expressed as the mean $\pm$ s.e.m. for six replicate wells (three replicates for thymus graft responders, animal 3) *P* values are given for the response to graft-type and third-party stimulators as compared with the response to host-type stimulators. The results indicate that whilst there is a minimal proliferative response to host-type stimulators, there is a significant response to both graft-type and third-party stimulators by cells from both host spleen and thymus grafts. Note that the response given by thymus graft cells, as in the case of normal thymus, is of a lower order than that given by spleen cells, suggesting that there is no graft accumulation of mature host peripheral cells reactive to graft type.

have found recently that chimaeric thymus lobes produced *in vitro* by recolonization of deoxyguanosine-treated lobes with allogeneic embryonic precursors, where there is no possible involvement of extrinsic mature cells, also contain cells capable of response to epithelial type stimulators in MLC (J.J.T.O., E.J.J. and R.K., unpublished observations).

We suggest that the most likely explanation for the growth of deoxyguanosine-treated thymus lobes in allogeneic recipients depends upon the elimination of dendritic and/or lymphoid cells and their precursors. We have found that deoxyguanosine-treated thymic epithelium is populated *in vitro* by donor dendritic cell precursors during the process of recolonization¹¹. It seems likely that treated thymic grafts are also populated by host dendritic cells *in vivo*, but remarkably these cells do not present the class II antigens of the donor epithelium to the host's immune system. The significance of class II antigen expression by epithelial cells remains unresolved but might be related more to the acquisition of MHC restriction by T cells than to tolerance induction.

We thank Alison Kingman for assistance with the photography and Lyndsey Halliwell for help in the preparation of the manuscript. The work was carried out with support from a MRC project grant to E.J.J. and J.J.T.O.

Received 23 February; accepted 18 April 1984.

1. Klein, J. *Science* **203**, 516–521 (1979).
2. Burnet, F. M. *Br. med. J.* **2**, 807–811 (1962).
3. Jenkinson, E. J., van Ewijk, W. & Owen, J. J. T. *J. exp. Med.* **153**, 280–292 (1980).
4. Cohen, A., Lee, J. W., Dosch, H. M. & Gelfand, J. *J. Immun.* **125**, 1578–1582 (1980).
5. Jenkinson, E. J., Franchi, L. F., Kingston, R. & Owen, J. J. T. *Eur. J. Immun.* **12**, 583–587 (1982).
6. Rouse, R. V., van Ewijk, W., Jones, P. L. & Weissman, I. L. *J. Immun.* **122**, 2508–2515 (1979).
7. Haynes, B. F., Shimizu, K. & Eisenbarth, G. S. *J. Clin. Invest.* **71**, 9–14 (1983).
8. Miller, J. F. A. P. *Thymus* **1**, 3–25 (1979).
9. Lafferty, K. J. *Transplantation* **29**, 179–182 (1980).
10. Barclay, A. N. & Mayrhofer, G. *J. exp. Med.* **153**, 1660–1671 (1981).
11. Owen, J. J. T. & Jenkinson, E. J. *Am. J. Anat.* (in the press).
12. Knight, S. C. *J. Immun. Meth.* **50**, R51–R56 (1982).
13. Czitrom, A. A. *et al. J. Immun.* **130**, 546–550 (1983).

Structural relationship of the SB β -chain gene to HLA-D-region genes and murine I-region genes

Eric O. Long*, Jack Gorski & Bernard Mach

Department of Microbiology, University of Geneva Medical School, 1205 Geneva, Switzerland

The major histocompatibility complex (MHC) regulates several aspects of the immune response. Class II antigens of the MHC control cellular interactions between lymphocytes¹. In man, at least three class II antigens (DR, DC and SB), consisting of distinct α - and β -chains, are encoded in the HLA complex². Sequence analysis has established that the DR and DC antigens are the respective structural counterparts of the murine I-E and I-A antigens². Molecular cloning of the SB β -chain gene has now enabled us to define its relationship to other class II genes. The DR, DC and SB β genes have diverged from each other to the same extent. In murine DNA and in cloned genes from the I region, the best hybridization of SB β DNA is with the E β 2 sequence. E β 2 may belong to a complete gene (E' β) because first domain sequences were found adjacent to it.

Limited amino acid sequence data have been reported for SB antigens³ and the extent of structural homology to the other class II antigens was unknown. Furthermore, it was not known whether a murine counterpart to the SB antigen existed but remained undetected, whether it existed but was lost during evolution or whether SB was generated by gene duplications after man-mouse divergence. In order to investigate these

Table 1 Nucleotide sequence homologies between human and murine class II β genes

	β -Chain domain											
	First				Second				Third			
	DR	DC	I-E	I-A	DR	DC	I-E	I-A	DR	DC	I-E	I-A
SB	76	77	67	72	77	79	77	76	68	71	71	70
DR		75	79	72		74	83	71		78	87	75
DC			68	79			74	83			75	86
I-E				74				70				70

The nucleotide sequences have been compared separately for each exon encoding the first, second and third domains of the β -chain. The fourth and fifth domains are too short to yield meaningful numbers. The numbers represent the homology between two sequences, expressed as a percentage. References for the human sequences are given in Fig. 1. The sequenced I-E β gene⁹ was from the H-2^d haplotype and the sequenced I-A β gene⁸ was from the H-2^b haplotype.

alternatives and to define the structure of the SB β -chain, we have isolated cDNA and genomic DNA clones for this chain. The cDNA clone SB β was obtained from a library, constructed from mRNA of a human B-cell line⁴, which was screened in conditions of relaxed stringency using DNA probes corresponding to the β -chains of DR and DC antigens. The purified insert from that cDNA clone was used to identify a corresponding gene from a cosmid library⁵. The complete DNA sequence of the cDNA clone, the first domain exon of the genomic clone and a map of the region comprising the SB α and β genes are presented elsewhere⁶. The deduced amino acid sequence (Fig. 1) matches the partial NH₂-terminal sequence of an SB3 β -chain³ in six out of seven positions. This single difference (amino acid 9 is tyrosine in SB3 and phenylalanine in our cloned gene) is best explained as an allelic difference because the genomic DNA library was constructed from DNA of an SB4 homozygous cell line. A cDNA clone derived from mRNA of another B-cell line was recently isolated and sequences related to it mapped in the SB subregion of HLA⁶. The amino acid sequence deduced from that cDNA clone shows 97% homology with the SB β -chain sequence presented here (Fig. 1).

Hybridization of the cDNA clone SB β to human genomic DNA digested with *Eco*RI revealed the presence of at least three bands (Fig. 2). The strongest band after high-stringency washes is a 10.5-kilobase (kb) fragment which corresponds to the SB β gene cloned in cosmid SB4 (ref. 5). Therefore, the cDNA clone SB β and the cosmid clone SB4 probably represent alleles of the same gene. Two additional fragments (7.5 and 3.6 kb) also hybridize well at a stringency which does not allow hybridization to DR or DC β genes; these bands represent at least one more SB-like β -gene sequence.

The overall structure of the SB β -chain is very similar to that of other class II β -chains (Fig. 1). Both extracellular domains contain disulphide bridges and the putative carbohydrate attachment site is at the same position, on asparagine 19. There are also two major differences. First, the alignment with the DR and DC β -chains can only be maintained if a gap of two amino acids is introduced after position 23. Most probably, a deletion of six nucleotides has occurred during the evolution of the SB β gene. Second, the cytoplasmic portion of these three β -chains is completely different. At the amino acid level the homology between SB and DC β -chains (69%) is ~5% greater than that between SB and DR (64%). However, at the nucleotide level the SB β gene is not more closely related to the DC β gene than to the DR β gene (Table 1). The SB, DC and DR β genes have diverged from each other to approximately the same extent. The SB β gene is neither I-E-like nor I-A-like; it is therefore appropriate to divide the HLA-D region into three distinct subregions (DR, DC and SB).

The nucleotide sequence homology is greater between DR and I-E or between DC and I-A β genes than intraspecies homologies between genes of different subregions (Table 1). This implies that the gene duplication which generated the DR/I-E and DC/I-A β genes must have preceded mammalian radiation. Because we have shown that the SB β gene ancestor

* Present address: Laboratory of Immunogenetics, NIAID, NIH, Building 5, Room B1-04, Bethesda, Maryland 20205, USA.

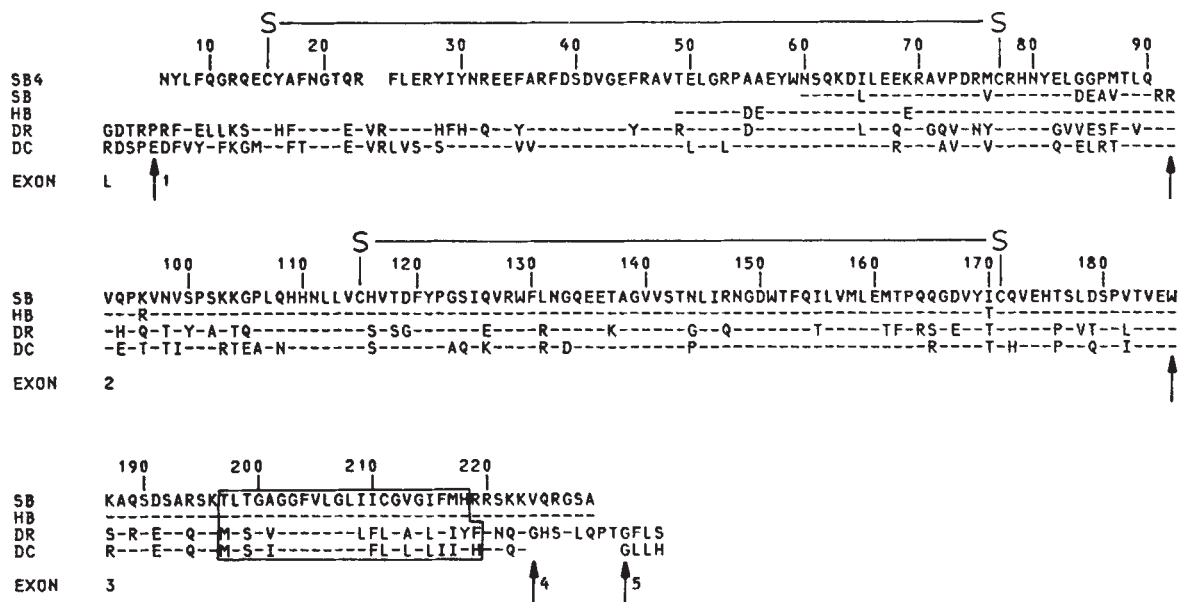


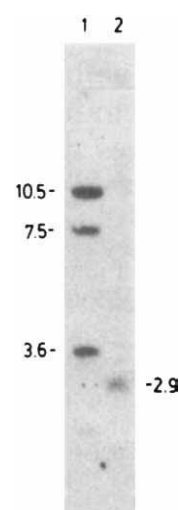
Fig. 1 Amino acid sequence comparison between SB, DR and DC β -chains. Amino acids are numbered based on the sequence of the SB β -chain. Hyphens indicate identity with the SB β sequence. Vertical arrows indicate exon-intron boundaries¹⁰. The hydrophobic transmembrane region is boxed. The sequences are derived from the following sources: SB4, genomic DNA clone encoding an SB4 β -chain⁵; SB, cDNA clone SB β (this paper); HB, cDNA clone HB β (ref. 6); DR, cDNA clone DR β I (ref. 4); DC, cDNA clone pII- β -1 (ref. 11). The cDNA clone SB β was isolated from a cDNA library constructed from mRNA of an HLA-heterozygous B-cell line⁴. The library was screened by colony hybridization¹² using inserts purified from the DR β -1 and the DC β -1 cDNA clones¹³ and with a final wash at 65 °C for 30 min in 0.15 M NaCl, 0.015 M trisodium citrate. Out of 37 positive clones identified in the screening, SB β was the only clone that had a weak homology with both the DR and the DC β probe. The purified insert from SB β was used to identify a genomic DNA fragment cloned in a cosmid vector⁵. The cosmid library was constructed from DNA of the consanguineous HLA-homozygous B-cell line HHK which types as DR6 and SB4. The nucleotide sequence of the cDNA clone SB β and the first domain of the SB4 β gene is presented elsewhere⁵.

is itself as 'old' as the DR-DC β gene duplication, its existence must also have preceded man-mouse divergence. We have investigated whether it is still present in mice today by hybridizing the cDNA clone SB β to an *Eco*RI digest of mouse (BALB/c) DNA (Fig. 2). After high-stringency washes a fragment of 2.9 kb was detected. In these conditions, the probe did not hybridize to the E β or A β genes, which are on fragments of 12 and 5.6 kb, respectively⁷. The SB β gene also hybridized better to this murine 2.9-kb *Eco*RI fragment than to the DR or DC β genes under the stringency used in this experiment (Fig. 2).

Most of the I region from BALB/c mice has been cloned recently in overlapping cosmids⁷. We used the cosmids carrying the A β and E β genes, as well as a sequence called E β 2 (which did not hybridize with β -chain first domain sequences) to test their degree of homology with the human β genes and to search for the 2.9-kb *Eco*RI fragment homologous to the SB β gene. The A β 2 sequence, which shows only 60% homology with other murine and human β genes⁸, was not included in our analysis. With human DNA probes corresponding mostly to the second and third domains of β -chains, the expected correlations DR/I-E and DC/I-A were observed (Fig. 3). The DR β probe also hybridized efficiently to E β 2. The best hybridization with SB β was E β 2, which is present on a 2.9-kb *Eco*RI fragment. The stringencies of the hybridization and washes (about 18 °C below the T_m of a DNA homoduplex) which allow a selective hybridization of SB β to E β 2 were the same as those used for the genomic blot shown in Fig. 2. We conclude that the sequence in the mouse genome which hybridizes best with the SB β gene is E β 2.

The same cross-hybridizations were carried out with DNA probes specific for the first domain of DR, DC and SB β -chains (Fig. 3). At a stringency which revealed the DR/I-E and DC/I-A correlations in the second and third domains, the first domains of the three human β genes hybridized about equally well to both A β and E β . This result is, in fact, expected, based on the analysis of nucleotide sequences. The central portion of the first domains has diverged more freely and the DR/I-E, DC/I-A

Fig. 2 Hybridization of a cDNA sequence for the SB β -chain with human (lane 1) and murine (lane 2) DNA. High-molecular weight DNA from the human B-cell line HHK (homozygous for HLA-DR6, SB4 by consanguinity) and from murine embryos of the strain BALB/c (*H*-2^d) was prepared as described previously¹⁴. 20 μ g of each were digested with the restriction endonuclease *Eco*RI (Boehringer Mannheim), electrophoresed in a 0.6% agarose gel and transferred to a nitrocellulose filter (Genatran 45; Plasco, Woburn, Massachusetts). The filter was hybridized with plasmid SB β which had been labelled with ³²P by nick-translation¹⁵ and washed exactly as described previously¹⁶ except that the final wash was in 75 mM NaCl, 7.5 mM trisodium citrate at 65 °C for 30 min. The size of the fragments is indicated in kilobases.



correlations are much weaker than in the second or third domains. The DR β probe also hybridized to a 13-kb fragment containing the leader exon of E β . A surprising result was the detection of a new fragment of 2.4 kb (fragment X in Fig. 3), distinct from the fragments containing the E β and A β genes. This fragment hybridized with all three human β -chain probes and was more apparent when hybridization stringencies were relaxed. Additional hybridizations with *Bam*HI-, *Hind*III- and *Kpn*I-digested cosmid 8.4 (data not shown) unambiguously mapped the hybridizing fragment as being adjacent and centromeric to E β 2. This suggests the existence of a novel mouse β -chain first domain (E β 1) near E β 2. Sequencing of this new gene (E' β) as well as expression studies are necessary to test whether it is complete and functional. The nucleotide sequence of E β should also clarify its evolutionary relationship to the SB β gene.

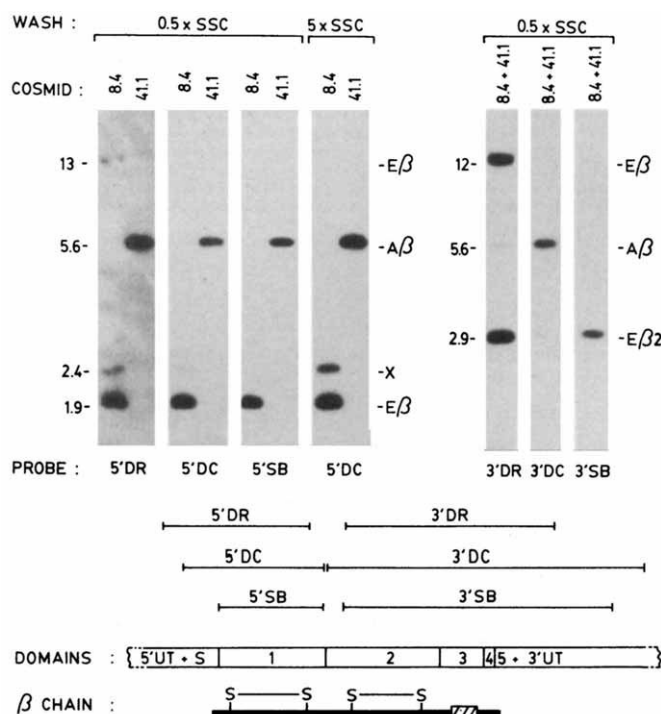


Fig. 3 Hybridization of HLA class II β -chain sequences with H-2 I-region genes. The cosmids 8.4 (carrying the E β and E β 2 genes) and 41.1 (carrying the A β gene), both derived from BALB/c DNA⁷, were digested with *Eco*RI. After separation by gel electrophoresis and transfer to nitrocellulose filters, the murine DNA fragments were hybridized with labelled DNA probes derived from human class II β genes. The sizes of the various fragments are indicated in kilobases on the left. The murine gene present on each fragment is indicated on the right. The human class II DNA probes were derived from the cDNA clones DR β -2 (ref. 13), DC β -1 (ref. 13) and SB β , and from a cosmid clone carrying an SB β gene⁵. The probe used for each filter is indicated, together with a map showing the position of each human DNA fragment relative to the coding sequence. The filters were hybridized at 65 °C for 14 h in 0.75 M NaCl, 50 mM sodium phosphate (pH 6.8), 5 mM EDTA, 0.02% Ficoll 400, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin, 50 μ g ml⁻¹ poly(G), 0.1% SDS and 2×10^5 c.p.m. ml⁻¹ of denatured, nick-translated¹⁵ DNA fragments. The filters were washed as described previously¹⁶ except for the salt concentration in the final wash, which was either 75 mM NaCl, 7.5 mM trisodium citrate (0.5 $\times$ SSC) or 0.75 M NaCl, 75 mM trisodium citrate (5 $\times$ SSC), as indicated.

We thank G. Mottet, C. Mattman and B. Grubenman for technical assistance, Drs M. Steinmetz for the murine I-region cosmids, M. Giphart for the B-cell line HHK and G. Shaw for typing the manuscript. This work was supported by the Swiss National Fund for Scientific Research.

Note added in proof: The preferential hybridization of SB β cDNA to the murine E β 2 sequence has been reproduced in another study (M. Rogers and D. Singer, personal communication). However, the nucleotide sequence of E β 2 shows that SB β is about equally related to E β , A β and E β 2 (N. Braunstein and R. Germain, personal communication).

Received 21 December 1983; accepted 4 May 1984.

- Klein, J., Juretic, A., Baxevanis, C. N. & Nagy, Z. A. *Nature* **291**, 455–460 (1981).
- Kaufman, J. F., Auffray, C., Korman, A. J., Shackelford, D. A. & Strominger, J. L. *Cell* **36**, 1–13 (1984).
- Hurley, C. K., Shaw, S., Nadler, L., Schlossman, S. & Capra, J. D. *J. exp. Med.* **156**, 1557–1562 (1982).
- Long, E. O., Wake, C. T., Gorski, J. & Mach, B. *EMBO J.* **2**, 389–394 (1983).
- Gorski, J., Rollini, P., Long, E. O. & Mach, B. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Roux-Dosseto, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6036–6040 (1983).
- Steinmetz, M. et al. *Nature* **300**, 35–42 (1982).
- Larhammar, D. et al. *Cell* **34**, 179–188 (1983).

- Saito, H., Maki, R. A., Clayton, L. K. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5520–5524 (1983).
- Larhammar, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **80**, 7313–7317 (1983).
- Larhammar, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3687–3691 (1982).
- Grunstein, M. & Hogness, D. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3691–3965 (1975).
- Long, E. O. et al. *Proc. natn. Acad. Sci. U.S.A.* **79**, 7465–7469 (1982).
- Gross-Bellard, M., Oudet, P. & Chambon, P. *Eur. J. Biochem.* **36**, 32–38 (1973).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
- Wake, C. T., Long, E. O. & Mach, B. *Nature* **300**, 372–374 (1982).

Structural and evolutionary analysis of HLA-D-region products

P. Travers[†], T. L. Blundell[‡], M. J. E. Sternberg^{‡§} & W. F. Bodmer^{*}

^{*} Imperial Cancer Research Fund, Lincoln's Inn Fields, London WC2A 3PX, UK

[‡] Laboratory of Molecular Biology, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, UK

[§] Laboratory of Molecular Biophysics, Department of Zoology, South Parks Road, Oxford OX1 3PS, UK

The major histocompatibility complex (MHC)—HLA in man and H-2 in mouse—encodes two classes of cell-surface antigens involved in the immune response. The amino acid sequences have been determined for a number of these molecules^{1–11}. Class I antigens, typified by the HLA-ABC antigens, are composed of a 43,000-molecular weight (MW) glycosylated transmembrane polypeptide with three external domains (α 1, α 2 and α 3), of which the one nearest the membrane (α 3) is associated with a 12,000-MW nonglycosylated polypeptide, β 2-microglobulin. The HLA-D-region or class II antigens, DR, DC and SB, are composed of two glycosylated transmembrane polypeptides, of MWs 34,000 (α -chain) and 28,000 (β -chain). Both chains have two external domains which presumably associate with each other, α 2, β 2 being membrane proximal and α 1, β 1 N-terminal and membrane distal. All four membrane-proximal domains (class I α 3, β 2-microglobulin, class II α 2 and β 2) have amino acid sequences that show significant similarities with immunoglobulin constant-region domains^{3,6,9,12,13}. This, together with the similarly placed internal disulphide bonds, suggests they might have an immunoglobulin-like structure (Fig. 1). We have now used computer graphics techniques to predict a detailed three-dimensional structure for the membrane-proximal domains of the class II antigens (α 2 and β 2) based on the known coordinates of immunoglobulin constant domains (Fig. 2). The transmembrane regions of class II antigens have been modelled as two α -helices packed together. The proposed structure accounts for conservation of amino acids and leads to evolutionary predictions.

Figure 1 shows the amino acid sequences of various MHC membrane-proximal domains and the human immunoglobulin C λ domain¹⁴, aligned on the basis of the presence of invariant residues, the maintenance of the pattern of hydrophobic/hydrophilic residues in the seven β -strand regions and a secondary structure prediction^{15,16} for the DR α 2 and β 2 chains. The numbering of the C λ domain, preceded by L, is used for reference. The structure of immunoglobulin fragment Fab (NEW) at 2.0 Å resolution¹⁴ was used to model the conformations of the HLA-DR membrane-proximal domains (C λ for DR α , and C μ 1 for DR β). Interactive computer graphics (program FRODO¹⁷, written by T. A. Jones and modified for the Evans & Sutherland PS2 by I. J. Tickle) enabled residue substitution, deletion and insertion to be performed followed by adjustment to the main- and side-chain torsion angles to obtain a stereochemically allowed model.

The detailed model structure (Fig. 2) is generally more compact than that of immunoglobulin domains, as deletions have

[†] Present address: Department of Microbiology, Stanford University School of Medicine, Stanford, California 94305, USA.

Fig. 1 Alignment of MHC membrane-proximal domains with an immunoglobulin constant-region domain. The sequences were aligned by eye to optimize the alignment of structurally invariant residues in the immunoglobulin sequence. Sequences used were those of the HLA-DR (DR $\alpha 2$)⁹, HLA-DC (DC $\alpha 2$)¹⁰, H-2 I-A (I-A $\alpha 2$) and H-2 I-E (I-E $\alpha 2$)¹¹ α -chains, HLA-DR (DR $\beta 2$)^{5,7} and DC (DC $\beta 2$)⁶ β -chains, HLA-B7 (HLA-B7 $\alpha 3$)¹², a possible HLA-ABC pseudogene (Ψ HLA-ABC $\alpha 3$)⁸, H-2K^b (ref. 4), β_2 -microglobulin (β_2m)³ and a human λ light chain (C_λ)¹⁴. The sequences are shown in the single-letter amino acid code. Underlining shows the positions of β -strands determined from the structure of Fab (NEW). The seven β -strands of the immunoglobulin constant region are lettered A-G (ref. 14).

DR $\alpha 2$	I T N V P E V T Y L T N S P V E L R E P N V L I C F I D K F T P P V Y N Y T W L R N Q K P V T T G V S E T V F L P R E D H L F R K F H Y L P F L P S T E D V Y D C R V E H W G L D E P L L K H W E F O A
DC $\alpha 2$	N E V P E V T Y F S K S P V T L G N P M T L I C L V D N I F P P V V N I T W L S N G S V T G G V S E T S F L S K S D S F F K I S Y L T F L P S A D E I Y D C K V E H W G L D E P L L K H W E F E I
I-A $\alpha 2$	N E A P Q A T V F F K S P V L L G O P N T L I C F V D N I F P P V V N I T W L R N S K S V T G G V S E T S F F V N R D S F H K L S Y L T F I P S D D I Y D C K V E H W G L E E P V L K H W E F E E
I-E $\alpha 2$	N V A P E V T Y L S R S P V N L G E P N L I C F I D K F S P P V Y N Y T W L R N R P V T E G V S E T V F L P R D D H L F R K F H Y L T F L P S T D D F Y D C E V D H W G L E E P L R K H W E F E I
DR $\beta 2$	R V Q P K V T Y P S K T O P L O H N H L L V C S V S G F Y P O S I E V R W F L N G E E A G G V S T G L I O D D W T F Q L V M L E T P R S G E V Y T C Q V E H P S V T S P L T V E W R A S
DC $\beta 2$	R V E P T V T I S P S R T E A L N H N H L L V C S V T D F Y P A Q I K V R W F R N D E T A A G G V S T P L I R N D W T F Q L V M L E M T P O R G D V Y T C H V E H P S L O S P I T V E W A O S
HLA B7 $\alpha 3$	R A D P P K T H Y T H N P I S D H E A T L R C W A L G F Y P A E I T L T W O R D G E D O T D T E L V E T R P A G D G T F O K W A A V V P S G E Q R Y T C H V O H E G L P K P L T L G W E P S S
Ψ HLA $\alpha 3$	R A D P P K T H M T H N P I S D H E A T L R C W A L G F Y P A E I T L T W O R D G E D O T D T E L V E T R P A G D G T F O K W A A V V P S G E Q R Y T C H V O H E G L P K P L T L R W E P S S
H-2K ^b	R T O S P K A N V T H N S R P D D K V T L R C W A L G F Y P A D I T L T W L N G E E L I O D M E L V E T R P A G D G T F O K W A S V V P L K G E V Y T C H V Y O O G L P O P L T L R W D E P P
β_2m	O R T P K I Q V Y S R H P A E N G K S N F L N C V Y S G F H P S D I E V D L L K N G E R I E K V E H S D L S F S K D W S F Y L L Y T F T P T K D E Y H C A V N H V T L S Q P K I V K W O R D M
C_λ	O P K A A P S V T L F P P S S E E L Q N K A T V C L I S D F Y P G A V T Y A N K A D S S P Y K A G V E T T T P S K O S N N K Y A A S S Y L T P E D N K S Y S C O V T H E G S T V E K T V A P T E C S

A B C D E F G

been placed in two back loops. The conformation around the tryptophan (L210) near the membrane, which is conserved in all MHC-encoded sequences, was adjusted so that this residue packs into the core of the domain, occupying the space left by the deletion of residues around L187. The HLA-D $\alpha 2$ domain is predicted to contain a salt bridge between the Glu L163 in the D-strand and Lys L176 in the E strand, as in the interaction between Glu L165 and Lys L176 in the membrane-proximal domain of the HLA-ABC chain. The DR β -chains and β_2 -microglobulin do not contain appropriate residues in these positions.

Both $\alpha 2$ and $\beta 2$ have a surface cluster of hydrophobic residues containing the conserved Phe L174 and Pro L143. In the α -chain, Trp L200 contributes to this cluster, while in the β -chain Trp L172 is involved. L168 is often proline and may also contribute to this grouping. The position of this hydrophobic cluster near the junction of the N-terminal and membrane-proximal domains suggests that it is involved in the interaction between these two domains, and that a similar hydrophobic patch should be expected on the surface of the N-terminal domain. Table 1 shows the residues involved in $\alpha 2$ - $\beta 2$ inter-domain contacts, together with the homologous residues from the HLA-DC, HLA-B7 and Fab (NEW) sequences and the main contact residues of the Mcg light chain dimer¹⁸. There is probably a hydrogen bond between Tyr L179 in the HLA-DR α -chain and Ser L137 in the HLA-DR β -chain, these residues being conserved in other class II α - and β -chains but not in either the HLA-ABC chains or in β_2 -microglobulin. As expected, the contacts between the HLA-DC membrane-proximal domains resemble those of HLA-DR but probably preclude heterodimeric combinations such as DR α with DC β . The contacts between HLA-ABC and β_2 -microglobulin are substantially different from those of the HLA-D-region products.

Table 1 Main contact residues involved in the interactions between the α and β membrane-proximal domains of the HLA-DR and -DC chains, between the HLA-ABC α -chain and β_2 -microglobulin and between the C_λ and C_H1 domains of Fab (NEW)

	DR		DC		HLA-B7		Fab(NEW)	
	α	β	α	β	α	β_2m	C_λ	C_H1
L118	T	T	T	T	T	Q	T	F
L120	→ L	Y	F	S	V	Y	F	L
L133	V	L	T	L	T	F	T	A
L135	→ I	V	I	V	R	N	V	C
L137	→ F	S	L	S	W	Y	L	L
L139	D	S	D	T	L	S	S	K
L162	S	S	S	S	E	H	E	H
L164	→ T	G	T	P	V	D	T	F
L175	R	Q	F	Q	Q	Y	A	S
L177	→ F	L	I	L	W	L	S	S
L179	→ Y	M	Y	M	A	Y	Y	V
L181	P	E	T	E	V	E	S	T

The main contact residues involved in the interactions between the two constant domains (C_L with C_H1) in the Mcg light-chain dimer, ($V_L C_L$)₂, are indicated by arrows. Numbering refers to the C_λ sequence of Fab (NEW).

The locations of residues conserved in all immunoglobulin constant and MHC-encoded membrane-proximal domains can be identified (Fig. 2a). One cluster is around the conserved cystine disulphide (Cys L136 and Cys L195) and involves Leu L134, Tyr L193 and Val L197. The only other conserved residue is Pro L143. If one excludes β_2 -microglobulin, however, Trp L150 also is conserved. Several residues are conserved between MHC sequences and not in C_λ domains: Asp L171, Phe L174, Pro L205 and Trp L210 (see above). Asp L171 and Phe L174 may be involved in the interaction between the membrane-proximal and membrane-distal domains; indeed the residue at L171 in immunoglobulin constant domains is involved in interactions with the variable domain.

The structural role of many residues conserved in the class II sequences can be identified (Fig. 2b). Leu L127, Trp L150 and Leu L180 pack into the hydrophobic core of the domain. The other residues often form hydrogen bonds (for example, Thr L118 and Asn L153). The residues common to the α -chains alone include those on the surface around Trp L120, the conformation and interactions of which are probably not found in immunoglobulins. Many of the other common residues (such as Pro L124, Val L125 and Pro L131) may be involved in interactions between the proximal domain and the membrane. There are also conserved surface clusters, for example Asp L194, Leu L151 and Thr L149. The region on the surface close to the connection to the distal domains also includes conserved residues, Pro L144, Val L145, Phe L166, Asn L112, Trp L200 and Gly L201.

The sequences of the transmembrane hydrophobic segments of the HLA-region products (Fig. 3), together with the predicted structure for DR α and β (Fig. 2d), were modelled in an α -helical conformation because this is a suitable structure for a single polypeptide chain that spans the lipid bilayer¹⁹ and because it allows all the hydrogen-bonding potential of the peptide groups to be fulfilled. In this conformation, both chains show faces of the α -helices almost entirely composed of glycines (see Fig. 2d legend). We suggest that the helices pack along these faces and that, as the side chain of glycine consists only of a hydrogen atom, the helices can lie parallel rather than at the more usual angles²⁰⁻²² dictated by interactions of larger side chains. Similar patterns of glycine residues can be seen in the internal fusogenic peptides of influenza virus haemagglutinin²³, Sendai virus F protein²⁴ and the transmembrane segment of glycoporphin²⁵. These peptides may also form close α -helix- α -helix interactions via glycine-rich faces. Comparison of the class II transmembrane peptides showed that the transmembrane segments are the most highly conserved parts of these molecules. For example, the homologous HLA-DC and H-2 I-A' α sequences share 21 out of 23 amino acids of the hydrophobic membrane-spanning peptide. However, for the HLA-ABC antigens, as β_2 -microglobulin has no transmembrane peptide, the conserved pattern of glycine residues does not occur, although some remnants of the pattern can be seen.

The α -chain connecting the membrane-proximal domains to the membrane-spanning domains is three residues longer than the corresponding β -chain sequence. As it contains three proline residues, it cannot adopt an extended conformation. Our model

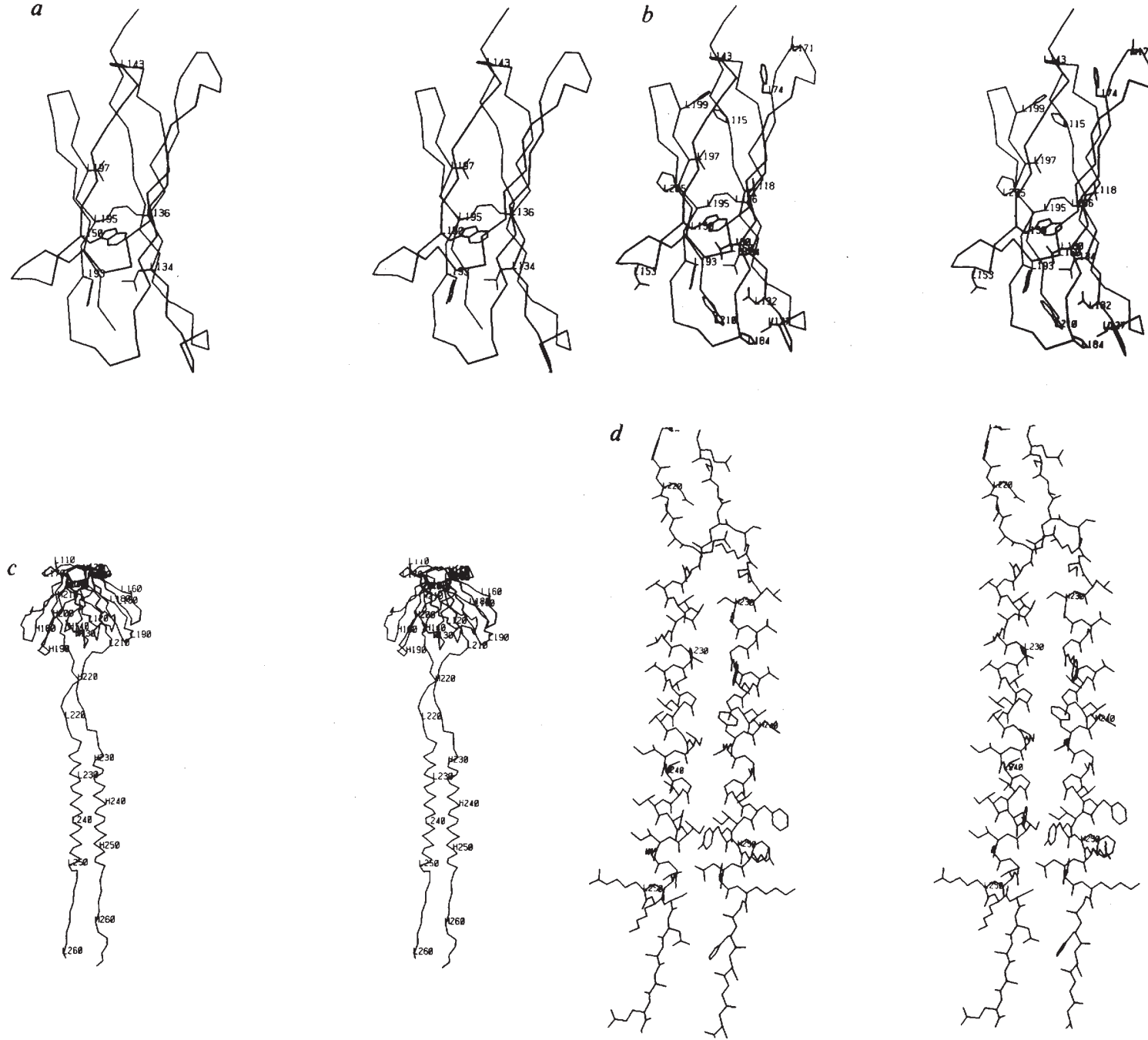


Fig. 2 Predicted structures for the HLA-DR membrane-proximal, transmembrane and intracellular regions. *a*, Stereo representation of HLA-DR proximal α -domain showing the path of the main chain from the connections between C α atoms and indicating side chains conserved between all immunoglobulin constant and MHC proximal domains. Trp L150, which is conserved except in β_2 -microglobulin, is also shown. *b*, As for *a*, but showing side chains conserved in all class II sequences. *c*, Stereo representation of the path of the main chain with the HLA-DR α -chain on the left and the β -chain on the right. The two membrane-proximal domains are shown at the top of the figure. The connecting peptides, transmembrane helices and intracellular peptides project from the base of the membrane-proximal domains. *d*, Predicted structure of the HLA-DR transmembrane region in more detail, showing the run of glycine residues in the two transmembrane helices. The α -chain lies on the left of the figure and the β -chain on the right. The predominantly hydrophobic helix is bounded on either side by charged residues. The sequence of glycine residues along each packing face is interrupted by Phe (or Leu in the DC and I-E α -chains), giving a sequence of residues along the helix of (Cys/Gly)-X-X-Gly-X-X-Gly-X-X-(Gly/Phe)-X-X-X-Gly-X-X-(Phe/Ile)- in the α - and β -chains. The Phe position in the α -chain is shifted relative to that of the β -chain by seven residues, or two helical turns, and this allows the α -helices to pack without close contact between the two Phe residues.

HLA-DR α	N V V C A L G L T V G L V G I I I G T I F I I K G V R K
HLA-DR β	M L S G I G G F V L G L I F L G L G L I I H R S Q K C L
HLA-A2	P T I P I V G I I A G L V L L G A V I T G A V V A A V M
Glycophorin	I G Y S I L L I T G I V G A M V G F I I L T I
Flu-HA2	G L F G A I A G F I E G G W E G M I D C W Y G F R
Sendai F	P F G A V I G I I A L G P A T

Fig. 3 Sequences of the HLA-DR α - and β -chain transmembrane peptides and of the HLA-B7 transmembrane peptide. The pattern of glycine residues in the HLA-DR sequences can readily be seen. Sequences of the transmembrane peptide of glycophorin²⁵, and the hydrophobic peptides of influenza haemagglutinin²³ and Sendai virus F protein²⁴ that are involved in promoting membrane fusion show a similar pattern of glycine residues that may also allow close interaction between helices.

proposes that these prolines introduce a twist in the α -chain which compensates for the difference in length of the two connecting peptides. The structures of the C-terminal intracellular peptides cannot be modelled with certainty. The DC β -chain is 8 amino acids shorter than the DR α -chain^{18,19} and contains only 7 amino acids from the end of the transmembrane segment predicted to traverse the 10 Å of the lipid head groups. It is possible, therefore, that only the terminal histidine residue of the DC β -chain is in the cytoplasm.

This modelling study of the conformation of the membrane-proximal and transmembrane regions of class II antigens has implications for the evolution of MHC-encoded and immunoglobulin molecules. Their common ancestor may have been a protein that had a structure superficially similar to that of class II products in possessing two closely packed transmembrane polypeptides²⁰, though these would be identical if the original molecule were a homodimer. These chains would have been organized into two domains, with the membrane-proximal domain having an immunoglobulin-like fold. Since immunoglobulins are found in primitive vertebrates (see ref. 26), the divergence of the MHC and immunoglobulin lineages probably occurred before the emergence of the vertebrates. This is consistent with a figure of 500–700 Myr for the divergence of the MHC products based on their sequence differences. At this time, before the development of a complex immune system, the functions of these products would not have been those now observed in mammals. The role of the ancestral MHC may therefore, as often suggested before, have been in cell-cell recognition, perhaps analogous to that which occurs, for example, during sponge aggregation²⁷. The alternative complement pathway also appears to have a counterpart in primitive invertebrates. Perhaps the ancestral MHC product interacted with it, but during the evolution of the MHC, this function was usurped by the immunoglobulins; this may explain the mapping of the C2, C4 and factor B complement within the MHC. In invertebrates, protection against invasion by foreign organisms is mediated by macrophage-like cells (coelomocytes, amoebocytes or archaeocytes). It is perhaps to these cells, and to a role in cellular recognition, that one should look for the precursors of the mammalian MHC.

We thank Drs J. Trowsdale, J. S. Lee, M. Owen, A. F. Williams and M. J. Crumpton for useful discussion; Drs J. Trowsdale, J. S. Lee, S. Weissman and E. Long for providing nucleotide sequences before publication; and the SERC for support of the interactive computer graphics workstation at Birkbeck College.

Received 12 October 1983; accepted 24 May 1984.

- Ploegh, H. L., Orr, H. T. & Strominger, J. L. *Cell* **24**, 287–299 (1981).
- Shackelford, D. A., Kaufman, J. F., Korman, A. J. & Strominger, J. L. *Immun. Rev.* **66**, 133–187 (1982).
- Smithies, O. & Poulik, M. D. *Science* **175**, 187–189 (1972).
- Coligan, J. E. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 3390–3394 (1978).
- Kratz, J. H. *et al. Hoppe-Seyler's Z. physiol. Chem.* **362**, 1665–1669 (1981).
- Larhammar, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 3687–3691 (1982).
- Long, E. O., Wake, C. T., Gorski, J. & Mach, B. *EMBO J.* **2**, 389–394 (1983).
- Malissen, M., Matissen, B. & Jordan, B. R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 893–897 (1982).
- Lee, J. S. *et al. Nature* **299**, 750–752 (1982).

- Auffray, C., Korman, A. J., Roux-Dosseto, M., Bono, R. & Strominger, J. L. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6337–6341 (1982).
- Benoist, C. O., Mathis, D. J., Kanter, M. R., Williams, V. E. & McDevitt, M. O. *Proc. natn. Acad. Sci. U.S.A.* **80**, 534–538 (1983).
- Orr, M. T., Lopez De Castro, J. A., Lancet, D. & Strominger, J. L. *Biochemistry* **18**, 5711–5719 (1979).
- Larhammar, D. *et al. Scand. J. Immun.* **14**, 617–622 (1981).
- Poljak, R. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **75**, 6002–6006 (1978).
- Chou, P. Y. & Fasman, G. D. *Adv. Enzym.* **47**, 45–148 (1978).
- Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97–120 (1978).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268–272 (1978).
- Edmundson, A. B., Ely, K. R., Abola, E. E., Schiffer, M. & Panagiotopoulos, N. *Biochemistry* **18**, 3953–3961 (1975).
- Tanford, C. *Science* **200**, 1012–1018 (1978).
- Rogers, J. *et al. Cell* **26**, 19–27 (1981).
- Crick, F. H. C. *Acta crystallogr.* **6**, 689–697 (1953).
- Chothia, C., Levitt, M. & Richardson, D. *J. molec. Biol.* **145**, 215–262 (1981).
- Gething, M. J., Bye, J., Skehel, J. & Waterfield, M. D. *Nature* **287**, 301–306 (1981).
- Gething, M. J., White, J. M. & Waterfield, M. D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2737–2740 (1978).
- Furthmayer, M., Galaray, R. E., Domita, M. & Marchesi, V. T. *Archs Biochem. Biophys.* **185**, 21–29 (1978).
- Hildemann, W. H. in *Comprehensive Immunogenetics* (eds Hildemann, W. H., Clark, E. A. & Raison, R. L.) 302–346 (Blackwell, Oxford, 1981).
- Humphreys, T. *Nature* **228**, 685–686 (1970).

High expression of cloned immunoglobulin κ gene in transgenic mice is restricted to B lymphocytes

Ursula Storb, Rebecca L. O'Brien,
Michael D. McMullen, Katherine A. Gollahan
& Ralph L. Brinster*

Department of Microbiology and Immunology, University of Washington, Seattle, Washington 98195, and
*Laboratory of Reproductive Physiology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

Immunoglobulin genes are normally expressed only in cells of the B lymphocyte lineage after a variable (V) and constant (C) gene rearrangement has occurred¹. To study the control of immunoglobulin gene expression in a defined situation, we have produced transgenic mice by microinjecting a rearranged mouse immunoglobulin κ gene (designated pB1-14) into fertilized mouse eggs². We present here the analysis of six different κ -transgenic mouse lines. All the transgenic mice express the microinjected κ gene in a completely tissue-specific fashion. Transcripts from pB1-14 are found at a high level in the spleen, but are undetectable in nonlymphoid tissues of testis, liver, kidney, heart, muscle, brain and thyroid gland. In lymphoid cell subpopulations, the level of pB1-14 transcripts is correlated with the relative number of B cells; there is no correlation with the proportion of T lymphocytes. We concluded, therefore, that the microinjected κ gene contains target sequences for B lymphocyte-specific gene activation signals that override the influence of the integration site.

The microinjected gene, pB1-14, encodes the κ chain produced by the myeloma MOPC-21 (ref. 3), and contains a variable region exon designated V κ M.21 and a constant region, C κ , exon. We present here the analysis of pB1-14 expression in different tissues of one original transgenic mouse and of first generation offspring from five other transgenic mice (see Table 1a for numbers, age and sex of mice). All the transgenic mice studied here contain multiple copies of the microinjected κ gene (Table 1a) which are inserted generally at a single site in tandem head-to-tail arrangement. As in one original transgenic mouse (no. 54) two different integration sites were seen which segregate in the offspring², this study deals with seven independent insertion sites.

All the transgenic mice and offspring which inherited the foreign gene have relatively large quantities of specific κ chains encoded by the microinjected gene in their blood serum as determined by two-dimensional gel electrophoresis (not shown). To determine the source of these proteins, the κ mRNA levels in different organs were analysed. RNA dot hybridization⁴ with

Table 1 Relative amounts of V_{κ} M.21 RNA in different transgenic and normal mice

(a) Survey of microinjected mice

Mice and organs	No. of copies of pB1-14 gene	pg κ RNA per μ g total RNA	% V_{κ} M.21
SB spleen (5 F; 4)	—	19	1.7
Old C57BL/6 spleen (2 F; 10)	—	18	0.67
No. 47 spleen (1 F, 2 M; 5.5)	24	254†	32
No. 48 spleen (5 F; 2)	48	4‡	39
No. 49 spleen (2 F, 1 M; 3)	16	27	40
No. 52 spleen (3 F, 2 M; 3)	32	17	25
No. 54 spleen (3 F; 4)	64, 24§	27	28
No. 55 spleen (1 M; 7)	32	96	25.8
No. 55 kidney (1 M; 7)	32	24	25.7

(b) κ RNA in young thymus and spleen||

Mice and organs	pg κ RNA per μ g total RNA	% V_{κ} M.21	κ RNA thymus/spleen¶
48 + Spleen	17.8	20.5	
48 + Thymus	2.3	17.2	0.12
48 - Spleen	11.9	3.3	
48 - Thymus	2.4	2.4	0.22

The values in parentheses are number sex (F = female, M = male) and age (in months) of the mice. In all cases, except no. 55, the organs from several mice were pooled before RNA extraction; 47–54 are F_1 offspring of transgenic mice. No. 55 is a first generation transgenic mouse. In a previous publication² on three of the mouse lines they were named A (now no. 48), B (no. 49) and C (no. 54). SB, normal SJL/J and C57BL/6 mice. SB and transgenic mice same as in Fig. 1. The copy number of the microinjected gene in transgenic mice was estimated by dot hybridization of kidney DNA (47 to 54) or tail DNA (55). The copy numbers for mice 48, 49 and 54 are lower than previously published² (50, 20, 95 and 40, respectively). The previous estimates were based on comparison with a pBR322 standard, whereas the values listed above were derived by comparison of the quantity of the C_{κ} gene with that of normal diploid mouse DNA and thus may be more accurate. Approximate quantities of total κ RNA and V_{κ} M.21-containing RNA were determined from dot hybridizations in probe excess, relative to a cloned DNA standard. After exposure of dot hybridizations to X-ray film (not shown), dots were cut out and counted in a liquid scintillation counter.

* Per cent of total κ RNA containing the V_{κ} M.21 sequence.

† We are unable to explain the very high level of κ RNA in these mice. Three other line 47 mice were analysed individually at 3 weeks old and found to contain levels of κ RNA in the range that is normal for other mice.

‡ This value may be artificially low because the RNA was partially degraded.

§ The original transgenic mouse 54 has two insertion sites which segregate in offspring; both types of offspring transcribe the microinjected gene in the spleen, but not the liver². Of the three mice shown here, two had the low and one the high copy number.

|| We examined 5½-week-old females of transgenic line no. 48 (+) or normal littermates (–).

¶ Relative amount of total κ RNA in thymus compared with spleen.

probes for C_{κ} and V_{κ} M.21 (the sequence specific for the microinjected gene) shows that the spleens of all mice contain κ RNA and that in transgenic mice a large proportion of this RNA represents V_{κ} M.21 sequences (Fig. 1). Spleens of normal mice, while producing similar levels of C_{κ} , do not contain sufficient V_{κ} M.21 sequences to produce a signal on this autoradiograph. The approximate quantities of total κ RNA and V_{κ} M.21-containing κ RNA were determined by quantitative dot hybridization (Table 1a). In normal spleens, only 0.7–3.3% of the κ RNA was encoded by a V_{κ} M.21 gene. This correlates well with the fact that only 0.6–4% of the κ chains present in the serum of normal mice contain this V region⁵. In spleens of

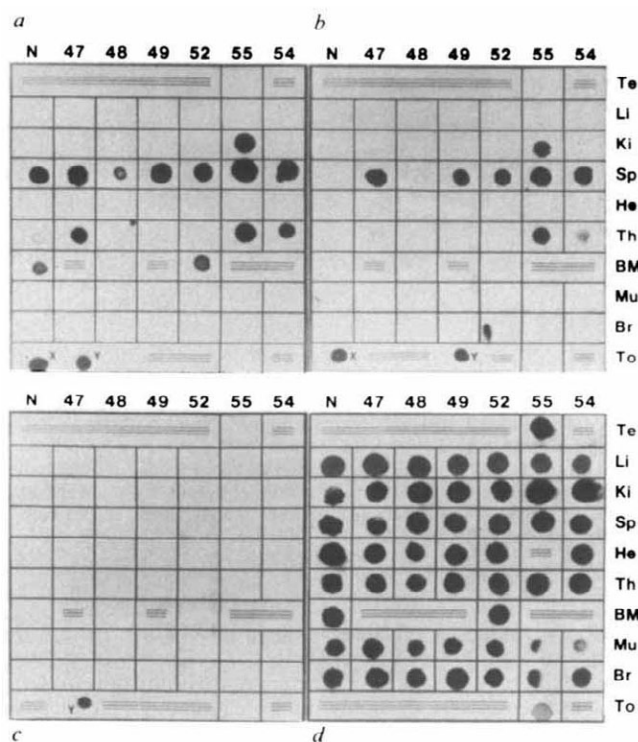


Fig. 1 Dot hybridization of RNA of various organs from transgenic and normal mice. Dot hybridizations were prepared as described elsewhere². Each dot comprised 30 μ g whole cell RNA denatured with glyoxal¹⁷. The description of the mice is given in Table 1. N, normal mice; Te, testis; Li, liver; Ki, kidney; Sp, spleen; He, heart; Th, thymus; BM, bone marrow; Mu, muscle; Br, brain; To, thyroid. X = 90 ng poly(A)⁺ RNA from myeloma MOPC-21. Y = 100 pg of unlabelled denatured probe DNA. =, No RNA added in this spot. The hybridization signals of spleen 48 RNA with M13 C_{κ} (a) and mpN2 (b) probes were low probably because of partial RNA degradation. In other RNA preparations (a total of 10 no. 48 mice) normal κ RNA levels were found (not shown). Small signals, which have been lost in the photographic reproduction, were present in the original X-ray films of thymuses of N, 48, 49 and 52 and bone marrow of 48, when hybridized with the M13 C_{κ} probe.

Methods: Hybridization conditions were 0.45 M NaCl, 0.045 M Na-citrate, 5× Denhardt's¹⁸ solution, 0.1% SDS, 50 μ g ml⁻¹ yeast tRNA (Sigma), 2.5 μ g ml⁻¹ (poly)A, 50 ng ml⁻¹ heat-denatured salmon testis DNA, 25 or 50 μ g ml⁻¹ ³²P-labelled nick-translated, heat-denatured cloned DNA probe (a–c), or cDNA from 700 ng ml⁻¹ liver poly(A)⁺ RNA (d), 10 ml total volume per nitrocellulose filter, at 65 °C for 24 h. The specific activity of the probes was 2–3×10⁸ c.p.m. μ g⁻¹. After hybridization, the filters were washed exhaustively in 30 mM NaCl, 3 mM Na-citrate, 0.05% SDS at 65 °C. The probes were M13 C_{κ} (a), containing ~500 nucleotides of C_{κ} and 3'-untranslated region²; and mpN2 (b), representing 287 nucleotides of the 3' end of the V_{κ} M.21 sequence. mpN2 was produced (B. Arp, unpublished) by subcloning into the *Hinc*II site of M.13mp7 *Alu*I-cut DNA of the phage M21B1 (ref. 3) containing the *Eco*RI insert of the MOPC-21 κ gene. M13 clones of *Alu*I-digested M21B1 DNA were screened with a short MOPC-21 V_{κ} probe, pES205 (ref. 19). The clone mpN2 was sequenced²⁰ and compared with the published sequence of V_{κ} M.21 (ref. 21) to determine that it contained no other insert of the κ gene or phage DNA (B. Arp, unpublished). The remaining probes were pBR322 (c), and cDNA of liver poly(A)⁺ RNA (d).

transgenic mice, however, 20.5–40% of the κ RNA was transcribed from the microinjected gene. There is no obvious correlation between the copy number of the microinjected gene and the relative amount of κ RNA encoded by that gene. The transcripts from the microinjected κ gene were of normal size in the spleens of all transgenic mice (not shown).

Some RNA preparations from thymus and bone marrow also contained κ RNA (see below). Nonlymphoid organs, however,

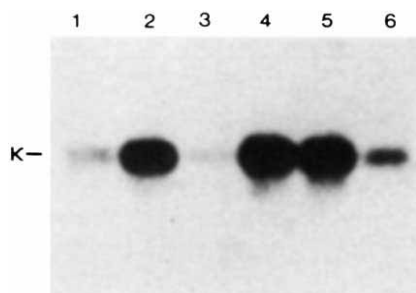


Fig. 2 Size of κ mRNAs in spleen cell populations. RNAs (2 μ g per slot, except in lane 1 where there was only a trace of RNA) from the experiment shown in Table 2 were treated with glyoxal¹⁷, electrophoresed in 1.5% agarose, 10 mM Na-phosphate pH 7.2, and blotted onto a nitrocellulose filter⁴. The blot was hybridized with the M13 C_{κ} probe as described in Fig. 1 legend. Lanes 1–3, normal littermates; 4–6, transgenic mice of line 48. 1, 4: total spleen cells; 2, 5: spleen cells released from anti-Ig; 3, 6: spleen cells passed through nylon wool. κ Indicates the position of the κ mRNA.

did not contain appreciable amounts of κ RNA (except the kidneys of no. 55—see below). Traces of κ RNA were apparent in nonlymphoid organs after long exposure of the dots to X-ray film, but these were probably due to circulating lymphocytes, as the levels in transgenic mice did not exceed those in normal mice (not shown). All the RNAs hybridized well with cDNA transcribed from liver poly(A)⁺ RNA (Fig. 1d). Therefore, the absence of hybridization of nonlymphoid RNA with κ probes truly reflects the absence, rather than degradation, of κ gene transcripts.

In contrast to the other transgenic mice, mouse no. 55 showed a large amount of $V_{\kappa}M.21$ RNA in the kidneys. Two alternative explanations for this anomaly were considered: the κ RNA could have been derived from infiltration by lymphoid cells due to an unknown cause, or the microinjected gene could have been activated in kidney cells proper. As this mouse was sterile, we could not study its offspring, so we used the RNAs in an attempt to distinguish between the two possibilities. One would expect that, if the microinjected gene was activated in kidney cells proper, endogenous κ genes would remain inactive. Thus, the ratio of $V_{\kappa}M.21$ -containing κ RNA to total κ RNA should be 1.0 if the RNA were derived from activation of the microinjected gene in kidney cells. The ratio should be lower and approximately the same as in spleen, if the RNA resulted from lymphoid cell infiltration. This was indeed the case—Table 1a shows that in kidney, just as in spleen, only about 26% of the κ RNA was encoded by the microinjected gene.

In all the transgenic mice, the microinjected pB1-14 gene is therefore expressed only in lymphoid cells. Next, it was important to determine whether the gene is expressed normally within the lymphoid cell lineages. Normally, only B lymphocytes, rather than T lymphocytes, express immunoglobulin genes at a high rate. Considerable amounts of κ RNA were found in the thymuses of some of the transgenic mice (lines 47, 54 and 55; Fig. 1). These transgenic mice were, however, relatively old, at an age when the thymus has regressed in size, and is difficult to distinguish from hilar lymph nodes: these are rich in B lymphocytes and could be the source of κ RNA in thymus preparations.

To obtain clean thymuses, transgenic mice and normal littermates of line no. 48 were killed at 6 weeks old, when the thymus is large and easily separable from lymph nodes. The RNAs from thymus and spleen were then analysed by semiquantitative dot hybridization (Table 1b). The thymuses of both transgenic and normal mice were found to contain some κ RNA. In the transgenic thymuses, this amounted to 12%, and in the normal thymuses to 22% of the level in spleen. Thus, the microinjected gene is not expressed at a high level in thymus T lymphocytes.

Table 2 Separation of spleen B and T cells

	Cell sample	% B cells	% T cells	c.p.m. hybridized with	
				C_{κ}	$V_{\kappa}M.21$
a	Normal total	43 (1.00)	24 (1.00)	873 (1.00)	0
b	Released from anti-Ig	30 (0.69)	24 (1.00)	505 (0.58)	0
c	After nylon wool	1 (0.02)	86 (3.86)	29 (0.03)	0
a'	Transgenic/total	46 (1.00)	32 (1.00)	1,612 (1.00)	469 (1.00)
b'	Released from anti-Ig	22 (0.47)	45 (1.38)	866 (0.54)	268 (0.54)
c'	After nylon wool	1 (0.02)	85 (2.63)	167 (0.10)	37 (0.08)

Spleen cells from two negative (normal) and two positive (transgenic) 5-week-old offspring of transgenic mouse no. 48 were washed, red cells were lysed and the spleen cells were added to a 10-cm plastic Petri dish coated with 5 mg ml⁻¹ rabbit anti-mouse immunoglobulin (Ig) and allowed to bind for 70 min at 4°C (ref. 15). Nonadhering cells were poured off and used to prepare T lymphocytes by using nylon wool columns¹⁶. Loosely adhering cells were washed off. Three cell pools [total spleen cells (a, a') spleen cells washed off the Petri dish (b, b') and cells which passed through nylon wool (c, c')] were analysed by RNA dot hybridization, RNA Northern blots, and fluorescence-activated cell sorting [Ortho; to determine the proportions of B and T lymphocytes after staining with fluorescein-conjugated rabbit anti-mouse immunoglobulin (Cappel) or after staining with fluorescein-conjugated anti-Thy 1 (Becton Dickinson)]. From each cell pool, total cellular RNA was prepared, spotted onto nitrocellulose filters (10 μ g for $V_{\kappa}M.21$, 2 μ g for C_{κ} probe), and hybridized in probe excess with mpN2 ($V_{\kappa}M.21$) and M13 C_{κ} . After exposure to X-ray film the hybrids were quantified by cutting the spots from the nitrocellulose filters and counting them in a scintillation counter relative to blank spots. The values in parentheses are relative numbers of cells or c.p.m. compared with the unfractionated spleen cells (a, a').

The low levels of κ RNA in transgenic and normal thymus may be derived from active B lymphocytes and plasma cells intimately associated with thymocytes⁶.

Although thymocytes, which are relatively immature T cells, did not seem to transcribe the microinjected gene, mature T cells might behave differently. To study the level of transcription in mature T cells, enriched populations were prepared from spleens. Table 2 shows that in both normal and transgenic spleens, with a decrease of B lymphocytes the levels of κ mRNA were coordinately decreased. Also, the cells expressing κ RNA encoded by the microinjected gene in transgenic mice ($V_{\kappa}M.21$ -positive RNA) were distributed in parallel with those expressing total κ RNA (Table 2). The decrease of κ RNA in the T-cell-enriched population was not due to degradation of the RNA as shown by the mature size of the RNA on Northern blots (Fig. 2). This experiment was repeated several times with similar results. When spleen B cells instead were enriched to a point where they represented >80% of the cell population, the $V_{\kappa}M.21$ -containing RNA was increased accordingly (not shown). It can be concluded that the microinjected κ gene is transcribed in B cells. The T cell-enriched population containing ~1% B cells, nonetheless still contained ~8% of the $V_{\kappa}M.21$ RNA; rather than representing a low level of expression of the microinjected gene in T cells, this probably results from copurification of plasma cells with T cells which increased from ~1% in the starting population to ~4% in the T cell-enriched population. As plasma cells produce high levels of κ RNA, it is likely that T cells do not express the microinjected κ gene at all.

Because rearranged immunoglobulin genes are normally only present in B lymphocytes, they are probably the only cells that contain the specific machinery for this type of gene shuffling. It is apparent from the findings in these transgenic mice, that the inability to rearrange is not the only factor which precludes immunoglobulin gene expression in other cell types. Although

B and T lymphocytes share many morphological and functional properties and are probably derived from a common stem cell⁷, only B lymphocytes seem to possess the signals which induce expression of the microinjected κ gene. It remains to be conclusively shown that these signals only appear in mature B cells; however, in the two transgenic lines for which bone marrow RNA was analysed, the total κ RNA level was not increased over that in normal bone marrow and V_{κ} M.21-containing RNA was not detectable (Fig. 1). As most bone marrow lymphocytes are pre-B cells⁸ it seems likely that pre-B cells do not transcribe the microinjected gene, that is, do not produce signals for κ gene expression.

We have not found any case of aberrant expression of the pB1-14 gene similar to that seen by Lacy *et al.*⁹ in transgenic mice carrying the rabbit β -globin gene. Statistically, this is not unexpected. Lacy *et al.*⁹ analysed about the same number of transgenic mice as we did and found only two cases of aberrant expression, skeletal muscle and testis, in separate mice (while none of the mice expressed the globin gene in erythrocytes). Their data, together with ours, suggest that the chance for integration into a genomic site which will be active in a particular tissue in an adult mouse is rather low. Given this background, it is striking that the pB1-14 κ gene is expressed in the spleen of all transgenic mice carrying this κ gene, that is, in all of seven different integrations. Integration of the microinjected pB1-14 gene is presumably random, as different 'junction fragments' appear on Southern blots of DNA from different transgenic lines (not shown). It seems highly improbable that the integration sites in these seven cases are normally active in B lymphocytes and inactive in other cells. More likely, the microinjected gene carries target sequences which allow its autonomous activation by B cell-specific signals. In other experiments, where transgenic mice lacked tissue-specific expression of a variety of microinjected genes¹⁰, the analogous target sequences may have been absent from the cloned genes. Experiments in search of the target sequences in the pB1-14 κ gene, which may include a putative 'enhancer'¹¹⁻¹⁴ region in the J_{κ} - C_{κ} intron, are now underway.

We thank Ben Arp for the production and sequencing of the mpN2 probe and the breeding of transgenic mice. This work was supported by grants from NIH (CA/AI 25754, DE01600, HD17321) and NSF (PCM81-07172). R.L.O'B. is supported by a NIH predoctoral training grant, K.A.G. by a postdoctoral fellowship from the Cancer Research Institute.

Note added in proof: In collaboration with O. Witte, K. Dennis and J. St Claire we have recently found that certain Abelson virus transformed pre-B-cell lines from transgenic mice carrying the pB1-14 κ gene produce mu-mRNA, but not the transgenic (or other) κ mRNA (unpublished data). This suggests that transcription signals for mu genes are different from and precede those for κ genes.

Received 3 April; accepted 25 April 1984.

- Hozumi, N. & Tonegawa, S. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3628-3632 (1976).
- Brinster, R. L. *et al. Nature* **306**, 332-336 (1983).
- Walfield, A. M., Storb, U., Selsing, E. & Zentgraf, H. *Nucleic Acids Res.* **20**, 4689-4707 (1980).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
- Rokhlina, O. V., Petrosyan, M. N., Ibragimov, A. R., Vengerova, T. I. & Bogachova, G. T. *Eur. J. Immun.* **13**, 397-403 (1983).
- Kemp, D. J., Wilson, A., Harris, A. W. & Shortman, K. *Nature* **286**, 168-170 (1980).
- Cooper, M. D., Kearny, J. & Scher, I. in *Fundamental Immunology* (ed. Paul, W. E.) 43-55 (Raven, New York, 1984).
- Butcher, E. C. & Weissman, I. L. in *Fundamental Immunology* (ed. Paul, W. E.) 109-127 (Raven, New York, 1984).
- Lacy, E., Roberts, S., Evans, E., Burtenshaw, M. D. & Constantini, F. D. *Cell* **34**, 343-358 (1983).
- Flavell, A. *Nature* **305**, 96-97 (1983).
- Chung, S.-Y., Folsom, V. & Wooley, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2427-2431 (1983).
- Emorine, L., Kuehl, M., Weir, L., Leder, P. & Max, E. E. *Nature* **304**, 447-449 (1983).
- Queen, C. & Baltimore, D. *Cell* **33**, 741-748 (1983).
- Picard, D. & Schaffner, W. *Nature* **307**, 80-82 (1984).
- Wysocki, L. J. & Sato, V. L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2844-2848 (1978).
- Henry, C. in *Selected Methods in Cellular Immunology* (eds Mishell, B. B. & Shiigi, S. M.) 182-185 (W. H. Freeman, San Francisco, 1980).
- McMaster, G. K. & Carmichael, G. G. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4835-4838 (1977).
- Denhardt, D. T. *Biochem. biophys. Res. Commun.* **23**, 641-652 (1966).
- Walfield, A., Selsing, E., Arp, B. & Storb, U. *Nucleic Acids Res.* **9**, 1101-1109 (1981).
- Messing, G., Crea, R. & Seeburg, P. H. *Nucleic Acids Res.* **9**, 309-321 (1981).
- Hamlyn, P. H., Gait, M. J. & Milstein, C. *Nucleic Acids Res.* **9**, 4485-4494 (1981).

Activated T lymphocytes produce a matrix-degrading heparan sulphate endoglycosidase

Yaakov Naparstek*†, Irun R. Cohen*§, Zvi Fuks‡ & Israel Vlodavsky‡

* Department of Cell Biology, The Weizmann Institute of Science, PO Box 26, Rehovot 76100, Israel

† Department of Medicine A, Hadassah University Hospital, Jerusalem, Israel

‡ Department of Radiation and Clinical Oncology, Hebrew University—Hadassah Medical School, PO Box 12000, Jerusalem 91120, Israel

We have previously found that lines of activated T lymphocytes specifically autosenitized to the basic protein of myelin (BP), on intravenous inoculation into syngeneic rats, were able to penetrate blood vessels, accumulate in the nervous system and cause experimental autoimmune encephalomyelitis (EAE)¹. An important question is how effector T cells reach such targets outside the walls of blood vessels. To investigate this we have studied *in vitro* the interaction of anti-BP effector T lymphocytes with the basement membrane-like extracellular matrix produced by vascular endothelial cells. We now report that activated but not resting T lymphocytes produce an endoglycosidase capable of degrading heparan sulphate side chains of the proteoglycan scaffold of the extracellular matrix. Moreover, the anti-BP T lymphocytes respond to BP presented by extracellular matrix by markedly enhanced elaboration of the endoglycosidase. These results suggest that tissue-specific antigens on blood vessel walls could direct lymphocyte homing by activating enzymes that facilitate penetration of the subendothelial basal lamina. They also suggest that effector T lymphocytes can recognize antigen which is not associated with a major histocompatibility complex signal.

The experimental system described here features two elements: anti-BP T lymphocytes and endothelial cell cultures producing extracellular matrix (ECM). Our method for raising antigen-specific lines of helper/delayed hypersensitivity lymphocytes has been described in detail². The lines were maintained by culture in growth medium lacking antigen but supplemented with a supernatant of cultures of lymphocytes stimulated by the mitogen concanavalin A (Con A) as a source of crude T-cell growth factor. Intravenous inoculation of Lewis rats with anti-BP line cells kept in growth medium had no discernible effect on the rats. However, after reactivation of the cells with BP in the presence of antigen-presenting cells syngeneic at the major histocompatibility complex (MHC), they proliferated vigorously and were able to enter and accumulate in the thymus and central nervous system, and induce EAE (refs 1-3). Activated anti-BP T lymphocytes blocked nerve conduction *in vitro*⁴. Moreover, after attenuation, activated anti-BP T lymphocytes could be used to vaccinate rats against active induction of EAE (refs 5, 6).

The bovine aortic endothelial cell system has been described elsewhere and shown to mimic the vascular endothelium *in vivo*^{7,8}. At confluence the cells form a monolayer composed of non-overlapping and closely apposed cells which express differentiated functions such as a non-thrombogenic apical surface and production of prostacyclin and anti-factor VIII antigen^{7,8}, but most pertinent here is a massive secretion of an underlying extracellular matrix in a polar fashion and similar in organization and supramolecular composition (types III, IV and V collagen, heparan sulphate, laminin, fibronectin) to naturally occurring basal laminae^{9,10}. The results of previous studies using high metastatic and low metastatic sublines of tumour cells¹¹⁻¹³ showed that higher amounts of heparan sulphate fragments released from the extracellular matrix by tumour cells correlated with greater potential to penetrate blood vessels and metastasize^{11,12}.

§ To whom correspondence should be addressed.

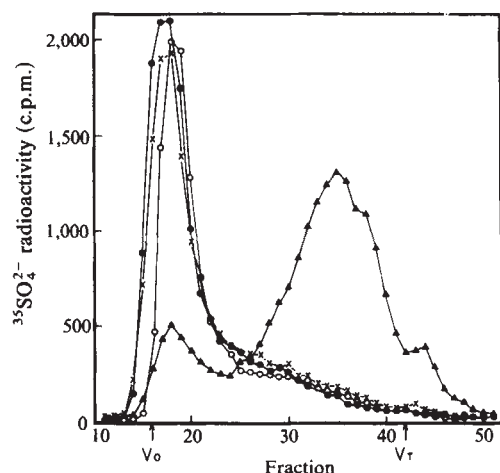


Fig. 1 Activated T lymphocytes degrade subendothelial extracellular matrix. Lewis rat anti-BP T lymphocytes were activated 72 h earlier by incubation with BP in the presence of irradiated syngeneic thymus accessory cells, or used without activation as described previously¹. 2×10^6 activated ($\blacktriangle$) or 4×10^6 non-activated ($\circ$) lymphocytes suspended in RPMI medium containing 10% fetal calf serum were then seeded on top of the labelled extracellular matrix. Labelled extracellular matrix was also incubated with medium containing irradiated thymus accessory cells and BP without lymphocytes ($\times$) or with activated lymphocytes in the presence of $10 \mu\text{g ml}^{-1}$ heparin ($\bullet$). The medium was collected after 24 h at 37°C and 0.5–0.7-ml aliquots of the centrifuged medium applied for gel filtration on Sepharose 6B columns (0.7×35 cm in most experiments and 1.1×70 cm in some experiments) as described previously¹¹. The excluded volume, V_0 , was marked by blue dextran, and the total included volume V_T , by phenol red. Similar elution profiles (K_{av} values) and recoveries were obtained whether the centrifuged media were subjected to gel filtration in dissociation conditions (4 M guanidine-HCl in 0.1 M sodium acetate, pH 5.5) or eluted with phosphate-buffered saline (PBS). Each experiment was performed at least four times and the variation in elution positions (K_{av} values) was less than 10%. Removal of any residual irradiated accessory cells by Ficoll-Hypaque density purification did not affect the results. The sulphate-containing moieties of glycosaminoglycans in the extracellular matrix were metabolically labelled by incubating subconfluent cultures of endothelial cells with radioactive ^{35}S as $\text{Na}_2[^{35}\text{S}]\text{O}_4$ which was taken up by the cells and incorporated into sulphated proteoglycans. The endothelial cells were then removed entirely by treatment with 0.5% Triton X-100, leaving the underlying, labelled extracellular matrix intact and firmly attached to the tissue culture plastic as described elsewhere^{9,10}. The activity of endoglycosidases in cells or media was assayed by incubating the labelled extracellular matrix with test material and measuring the release into the culture medium of cetylpyridinium chloride-precipitable, ^{35}S -labelled material eluted in the low-molecular weight ($K_{av} = 0.63$) region when analysed by Sepharose 6B gel filtration¹¹. This material contained mostly heparan sulphate sequences smaller than intact heparan sulphate side chains ($K_{av} = 0.32$) released by treating the extracellular matrix with papain or alkaline borohydride¹⁰.

Our strategy in the present study was to incubate labelled ECM with anti-BP T lymphocytes and assay the low-molecular weight heparan sulphate-containing fragments released into the culture medium as a measure of the degradation of the extracellular matrix by specific endoglycosidase. Figure 1 shows that non-activated anti-BP T lymphocytes maintained in growth medium had little or no enzyme activity detected during 24 h of incubation. Populations of irradiated accessory thymocytes from naive Lewis rats that included normal thymus lymphocytes and antigen-presenting cells also showed a very low specific enzyme activity. In contrast, anti-BP T lymphocytes that had been activated 72 h earlier by incubation with BP and irradiated (1,500 R) syngeneic thymus accessory cells as a source of antigen-presenting cells showed high endoglycosidase activity as indicated by the release of low-molecular weight ($K_{av} = 0.63$, $M_r 10^4$) labelled degradation products. The elaboration of endo-

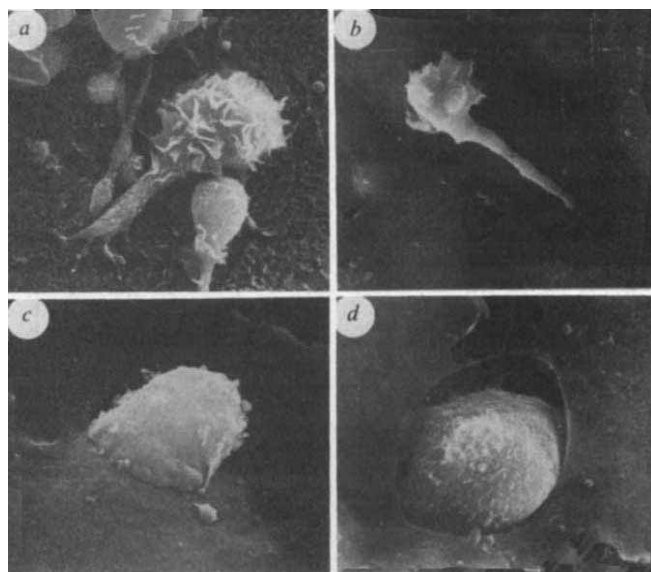


Fig. 2 Activated T lymphocytes invade cultured vascular endothelium. Activated anti-BP T lymphocytes¹ were separated by Ficoll density centrifugation from any contaminating accessory cells and incubated (2–6 h; 2×10^6 cells per 35-mm culture dish) on top of denuded extracellular matrix (a) or intact confluent monolayers of bovine vascular endothelial cells (b–d). The cultures were then washed to remove non-adhering cells, fixed (2.5% glutaraldehyde in PBS) and processed for scanning electron microscopy as described elsewhere¹³. a, Adherence and extension of pseudopods at 12 h incubation in direct contact with the extracellular matrix ($\times 1,920$). About 30% of the activated lymphocytes adhering to endothelial cells manifested within 3 h initial invasion by means of extension of pseudopods (b, $\times 1,560$) or penetration of the endothelial cell monolayers as seen after 6 h of incubation (c, $\times 3,600$; d, $\times 3,000$).

glycosidase activity was not limited to T lymphocytes reactive to BP and similar results were obtained with T lymphocyte lines reactive to ovalbumin or to the purified protein derivative of *Mycobacterium tuberculosis* (PPD). The anti-PPD line expressed endoglycosidase activity on incubation with PPD but not with BP (not shown).

To rule out the possible contribution of residual accessory cells surviving irradiation, Ficoll-Hypaque-purified activated lymphocytes were suspended in fresh medium, both in the absence or presence of 10% fetal calf serum and a similar degradation profile was obtained (not shown). To examine whether endoglycosidase was secreted, the activated lymphocytes were resuspended in serum-free medium, incubated (48 h, 37°C) in growth conditions and 1 ml of the conditioned medium incubated with the $\text{Na}_2[^{35}\text{S}]\text{O}_4$ -labelled ECM. More than 80% of the released radioactivity was in the form of heparan sulphate degradation products eluted as peak II, reflecting the presence of a soluble endoglycosidase similar to that of intact activated cells (not shown). The fact that this activity was obtained in the absence of serum indicates that the cells secreted the enzyme and did not merely activate an enzyme present in serum. That this activity was specific for heparan sulphate was suggested by its inhibition in the presence of $10 \mu\text{g ml}^{-1}$ heparin (Fig. 1). Previous studies have shown that more than 70% of the $^{35}\text{SO}_4^-$ radioactivity in the labelled extracellular matrix is incorporated into heparan sulphate^{11,12}. Up to 85% of this radioactivity was released in the form of degraded heparan sulphate side chains on incubation of the extracellular matrix with activated anti-BP cells. The low-molecular weight degradation products were confirmed to be fragments of heparan sulphate side chains by their sensitivity to deamination with nitrous acid¹⁴, precipitation with 0.05% cetylpyridinium chloride in 0.6 M NaCl (ref. 14), and resistance to digestion with papain

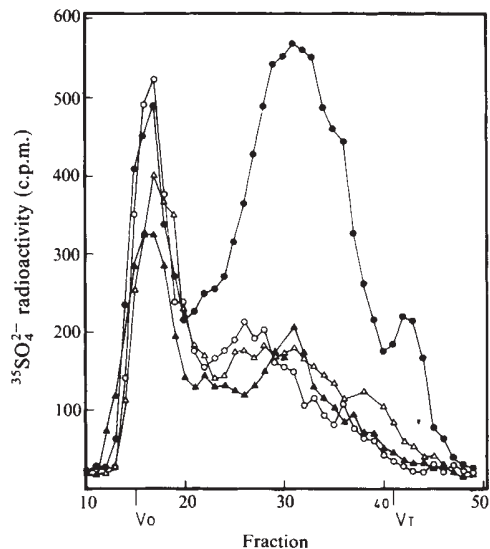


Fig. 3 Activated anti-BP T lymphocytes respond to BP-pulsed extracellular matrix by augmented endoglycosidase activity. Ficoll-purified activated (●, ○) or unactivated (▲, △) anti-BP T lymphocytes (4×10^6) were incubated for 12 h (activated lymphocytes) or 48 h (non-activated lymphocytes) with labelled extracellular matrix. Some of the extracellular matrix-coated dishes had been pulsed with BP ($250 \mu\text{g ml}^{-1}$ for 2 h at 37°C) and extensively washed with PBS before the addition of cells (●, ▲) and others were used without BP pulsing (○, △). The medium was then collected and analysed for specific endoglycosidase activity measured by the amount of radioactivity eluted as peak II as described in the legend to Fig. 1 and ref. 11.

and chondroitinase ABC. Thus, activation of anti-BP T lymphocytes was associated with induction of specific heparan sulphate endoglycosidase activity.

Figure 2 shows that activated anti-BP T lymphocytes could invade the vascular endothelium, primarily between adjacent endothelial cells adherent to the extracellular matrix. Invasion began by an extension of a cytoplasmic process (Fig. 2b) in about 30% of the cells, in a manner similar to that observed with metastatic lymphoma cells¹³. Non-activated line cells showed a significantly lower (<10%) rate of attachment and invasion through the endothelial cell monolayer. Hence, the production of extracellular matrix-degrading heparan sulphate endoglycosidase by anti-BP T lymphocytes and their ability to invade the vascular endothelium *in vitro* correlated with the capacity of the activated lymphocytes to accumulate in the central nervous system and induce EAE *in vivo*¹.

We also investigated whether the presence of specific antigen bound to the extracellular matrix influenced endoglycosidase activity. After incubation of extracellular matrix with BP, unattached BP was washed off, and then endoglycosidase activity of anti-BP T lymphocytes incubated on the BP-extracellular matrix complex was measured (Fig. 3). We found that activated anti-BP T lymphocytes responded to the BP-extracellular matrix complex by a markedly increased endoglycosidase activity relative to that observed when incubated with matrix alone for only 12 h, too short a time to detect appreciable enzyme activity produced by activated lymphocytes on untreated matrix. Non-activated anti-BP T lymphocytes showed little or no specific enzyme activity over 48 h of incubation whether or not the extracellular matrix was pulsed with BP (Fig. 3). In a series of four experiments, we estimated that the BP-extracellular matrix complex elevated the endoglycosidase activity expressed by the anti-BP lymphocytes by about fivefold. Thus, the presence of BP antigen bound to the ECM triggered augmented expression of the enzyme. Activated lines of T lymphocytes specific for purified protein derivative did not respond to the BP-extracellular matrix complex (not shown). Measurement of incorporation of ^3H -thymidine into DNA demonstrated that the BP-

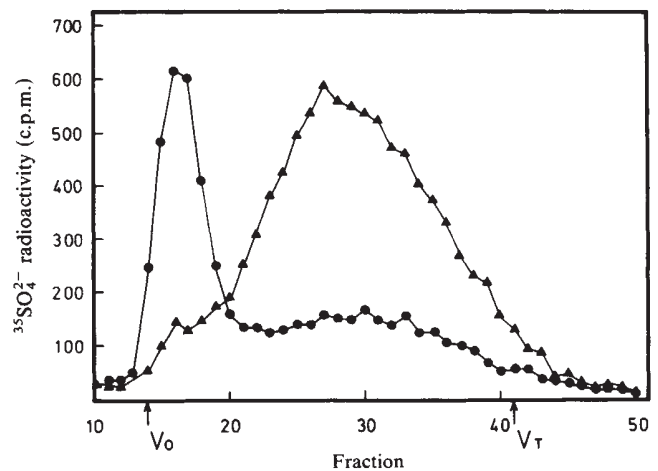


Fig. 4 Non-activated T lymphocytes respond to Con A-pulsed extracellular matrix by augmented endoglycosidase activity. Ficoll-purified non-activated anti-BP T lymphocytes (1×10^6) were incubated (48 h, 37°C) with untreated labelled extracellular matrix (●) or with labelled matrix that had been pulsed with $250 \mu\text{g ml}^{-1}$ Con A (Miles Yeda) for 2 h at 37°C , followed by five washes with PBS (▲). The medium was then centrifuged and analysed by gel filtration as described for Fig. 1.

extracellular matrix complex did not induce proliferation of anti-BP T lymphocytes (not shown).

Figure 4 shows that specific enzyme activity could be induced by incubating for 48 h non-activated anti-BP line lymphocytes with extracellular matrix that had been treated with the T lymphocyte mitogen Con A. Activated T lymphocytes responded to Con A-treated extracellular matrix within 12 h (not shown). The induction of enzyme activity by the Con A-extracellular matrix complex was not accompanied by proliferation of the line cells (not shown). We do not as yet know why BP-extracellular matrix induced augmented endoglycosidase activity only in activated anti-BP T lymphocytes, while Con A-extracellular matrix induced a high enzyme activity in either activated or non-activated anti-BP T lymphocytes. In any case, the absence of lymphocyte proliferation in the presence of BP or Con A was functional evidence of the absence of accessory cells that could have presented BP or Con A.

The present finding that rat T lymphocytes of the helper/delayed hypersensitivity class recognized antigen or mitogen associated with extracellular matrix of bovine origin in the absence of syngeneic MHC accessory cells indicates that association of antigen with MHC may not be a universal requirement for recognition by T lymphocytes, and implies that T lymphocytes might have receptors capable of recognizing and responding to antigen alone (see ref. 15). But it is clear that activated T lymphocytes, like metastatic tumour cells^{12,13}, have molecular machinery for penetrating the vascular basal lamina and that specific antigen leaking from the target and sequestered by the extracellular matrix may direct the expression of this potential.

We thank Mr H. Otmý for help in preparing the lymphocyte cultures, Ms R. I. Michaeli for technical assistance and Ms D. Gurfel for help with the scanning electron microscope. This work was supported by PHS grant CA 30289 to I.V. by the National Cancer Institute, by grant NS 18168 awarded to I.R.C. by the National Institute of Neurological and Communicative Disorders and Stroke, NIH and by the joint research fund of the Hebrew University and Hadassah awarded to Y.N. I.R.C. is the incumbent of the Mauerberger Professorial Chair in Immunology.

Received 25 November 1983; accepted 9 May 1984.

1. Naparstek, Y. *et al.* *Eur. J. Immun.* **13**, 418-423 (1983).
2. Ben-Nun, A. & Cohen, I. R. *J. Immun.* **129**, 303-308 (1982).
3. Naparstek, Y. *et al.* *Nature* **300**, 262-263 (1982).

4. Yarom, Y., Naparstek, Y., Holoshitz, J., Ben-Nun, A. & Cohen, I. R. *Nature* **303**, 246–247 (1983).
5. Ben-Nun, A., Wekerle, H. & Cohen, I. R. *Nature* **292**, 60–61 (1981).
6. Cohen, I. R., Ben-Nun, A., Holoshitz, J., Maron, R. & Zerubavel, R. *Immun. Today* **4**, 227–230 (1983).
7. Gospodarowicz, D., Moran, J., Braun, D. & Birdwell, C. *Proc. natn. Acad. Sci. U.S.A.* **73**, 4120–4124 (1976).
8. Vlodavsky, I. & Gospodarowicz, D. *J. J. supramolec. Struct.* **12**, 73–114 (1979).
9. Vlodavsky, I., Lui, G.-M. & Gospodarowicz, D. *Cell* **19**, 607–616 (1980).
10. Gospodarowicz, D. & Vlodavsky, I. *Proc. Cancer Res. Ther.* **22**, 73–104 (1982).
11. Vlodavsky, I., Fuks, Z., Bar-Ner, M., Ariav, Y. & Schirmacher, V. *Cancer Res.* **43**, 2704–2711 (1983).
12. Nakajima, M., Irimura, T., Di Ferrante, D. T., Di Ferrante, N. & Nicolson, G. L. *Science* **220**, 611–613 (1983).
13. Vlodavsky, I., Schirmacher, V., Ariav, Y. & Fuks, Z. *Inv. Metastasis* **3**, 81–97 (1983).
14. Roden, L., Baker, J. R., Citonelli, J. A. & Mathews, M. B. *Meih. Enzym.* **28**, 73–140 (1972).
15. Matzinger, P. *Nature* **292**, 497–501 (1981).

Autoimmune human T lymphocytes specific for acetylcholine receptor

Reinhard Hohlfeld*, Klaus V. Toyka*,
Kurt Heininger*, Hans Grosse-Wilde† & Inge Kalies‡

* Department of Neurology, University of Düsseldorf,
Moorenstrasse 5, 4000 Düsseldorf, FRG

† Division of Immunogenetics, University of Essen,
Virchowstrasse 171, 43 Essen, FRG

‡ Institute for Clinical Immunology, University of Erlangen,
Krankenhausstrasse 12, 852 Erlangen, FRG

Myasthenia gravis is one of the best characterized human autoimmune disorders^{1–4}. Circulating autoantibodies to the nicotinic acetylcholine receptor (AChR) at the neuromuscular junction play a prominent part in the effector phase of the disease^{5,6}, but little is known about the induction phase, that is, the immunoregulation. Indirect evidence, such as thymic abnormalities⁷ and the association with certain histocompatibility antigens (for example HLA-B8, DR3^{8,9}) suggests a defect of immunoregulation at the level of thymus-dependent (T) lymphocytes. We report here on the isolation of autoreactive T cells from six patients with myasthenia gravis. From one of these patients, who is homozygous for HLA-DR3, we established a long-term T-cell line. The line cells are specific for purified fish and human AChR, display the surface phenotype of inducer/helper T cells and are genetically restricted to HLA-DR3. AChR-induced proliferation could be inhibited with two monoclonal antibodies against monomorphic DR determinants and also with DR3-specific alloantisera.

It is known that peripheral blood mononuclear cells (PBM) from 50–70% of patients with myasthenia gravis, but not from healthy donors, proliferate weakly when cultured with AChR from fish electric organs^{10–12} or with human receptor¹³. As a prerequisite to the isolation of AChR-reactive T lymphocytes from the peripheral blood, we identified six 'AChR-responders' among our patients (see legends to Figs 1, 3). PBM from these patients were stimulated with *Torpedo* AChR or with tuberculin (PPD) in bulk cultures. The reactive lymphoblasts were then separated from the unstimulated cells by centrifugation on discontinuous Percoll density gradients and propagated and expanded in number in the presence of T-cell growth factor (interleukin-2). With this protocol it was possible to select immunospecific T-lymphocyte populations from the peripheral blood which were enriched several hundredfold in reactivity to the respective antigen (Fig. 1). Long-term cultures (>4 weeks) of selected lymphocytes required repeated restimulation with the respective antigen plus mitomycin C-treated PBM as a source of antigen-presenting cells. AChR-reactive T cells from one of the six patients (patient 1) could be expanded greatly in number and have been maintained in continuous culture for over 6 months as a long-term cell line. Several putative sub- 'clones', which were derived from single cells of the original line by limiting dilution¹⁴, had essentially the same characteristics in terms of AChR reactivity, genetic restriction and

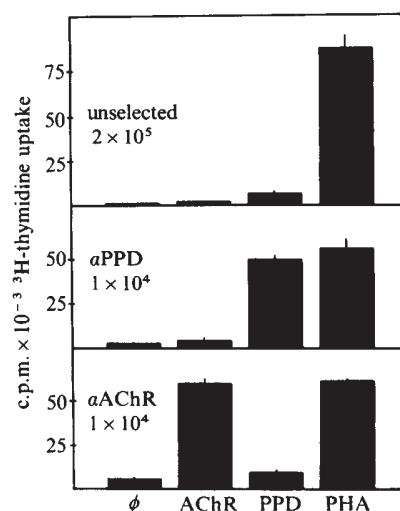


Fig. 1 Purification of AChR- or PPD (purified protein derivative of tuberculin)-specific T cells from the peripheral blood of a patient with myasthenia gravis who was reactive against AChR and PPD (patient 2; see below). Proliferative response (c.p.m. ± s.d.) in microproliferation assay of unselected PBM (top) or selected lymphocytes (middle and bottom). The antigen-induced response of 10^4 purified blast cells was tested after 5–10 days of propagation with interleukin-2 (see below) and was approximately 10 times higher than that of 2×10^5 unselected fresh PBM (enrichment factor 200). The corresponding initial enrichment factors of AChR-specific T cells in five additional patients were 160 (patient 1), 140 (patient 3), 80 (patient 4), 160 (patient 5), and 60 (patient 6). φ, No antigen added.

Methods: PBM were isolated by centrifugation on Ficoll–Hypaque (Pharmacia). For the microproliferation assay of unselected PBM, 2×10^5 cells were cultured in triplicate in round-bottom microtitre wells (Nunc) in 0.2 ml of RPMI 1640 (Gibco) supplemented with 10% inactivated (56 °C, 30 min) pooled human serum (or with 10% fetal calf serum (FCS) where indicated), 2 mM L-glutamine, 100 U per ml penicillin, 50 μg ml⁻¹ streptomycin (culture medium) and antigens, *Torpedo* AChR (5 μg ml⁻¹), PPD (Behringwerke; 1 μg ml⁻¹), or phytohaemagglutinin (PHA, Wellcome; 1 μg ml⁻¹). The purification of *Torpedo* (and human, see Fig. 2) AChR by chromatography on *Naja naja siamensis* α-neurotoxin affinity columns was performed as described previously^{22,26}. The specific activities were 5–7.5 (*Torpedo*) and 4 (human AChR) nmol α-bungarotoxin binding sites per mg protein. After 72 h the cultures were pulsed with ³H-thymidine (1 μCi per well, specific activity 5 Ci mmol⁻¹, Amersham) for 16 h. The cells were collected with a Titertek multiple harvester and thymidine incorporation was measured in a liquid scintillation counter. Antigen-specific cells were selected as follows: PBM were incubated in 10 ml plastic tubes in culture medium containing 5 μg ml⁻¹ *Torpedo* AChR or 1 μg ml⁻¹ PPD (4×10^6 cells per ml per tube) for 4 days. The lymphoblasts generated were isolated on a discontinuous Percoll (Pharmacia) density gradient²⁷. The blasts were seeded at 1×10^5 cells per ml per well in 24-well Costar culture plates in culture medium supplemented with 7.5% of a commercial interleukin-2-containing preparation (Lymphocult T, Biotest), that is, propagation medium. After 5–10 days of culture in propagation medium, the purified cells were tested for proliferative activity: 1×10^4 cells were cultured with or without 2×10^5 mitomycin C-treated PBM in the presence or absence of antigen as described for the microproliferation assay of unselected PBM.

phenotype (see below), indicating that the long-term cell line is probably composed of only a small number of diverse clones.

Several lines of evidence indicate that the long-term cell line represents purified autoreactive T lymphocytes. (1) The cells which were selected and propagated exclusively with *Torpedo* AChR also react with purified human AChR (Fig. 2). This suggests that the cell line is indeed reactive against self-AChR and not against some other component which might contaminate the purified AChR preparations. (2) The line is composed exclusively of T cells. This is reflected by the fact that autologous

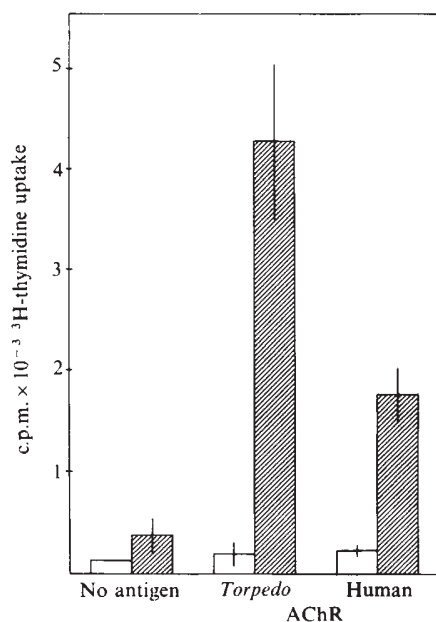


Fig. 2 Autoreactivity of the long-term T-cell line (patient 1). AChR response (c.p.m. $\pm$ s.d., triplicate cultures) in microproliferation assay (see Fig. 1) of 1×10^4 line cells after 12 weeks of continuous culture. Open bars: no antigen-presenting cells added. Hatched bars: 2×10^5 mitomycin C-treated autologous PBM added. The cell lines reacted with both *Torpedo* ($5 \mu\text{g ml}^{-1}$) and human ($5 \mu\text{g ml}^{-1}$) AChR (see legend to Fig. 1).

Methods: The long-term cell line was established as follows: Immunospecific, AChR-reactive T cells were selected with *Torpedo* AChR as described in Fig. 1. Every 4 days the cells were reseeded at 1×10^5 per ml in fresh propagation medium. Intermittently (every 8–12 days), cells were stimulated with $5 \mu\text{g ml}^{-1}$ *Torpedo* AChR plus 1×10^6 mitomycin C (Sigma)-treated ($50 \mu\text{g ml}^{-1}$; 30 min) syngeneic PBM per 10^5 blasts per ml of culture medium.

mitomycin C-treated PBM are required as a source of antigen-presenting accessory cells (Fig. 2). (3) Purity of the line was also established by its complete loss of alloreactivity which is a hallmark of unselected T-cell populations. Loss of alloreactivity made it possible to determine the genetic restriction of the long-term cell line and of purified AChR-reactive T cells from the five other patients in co-cultures with antigen-presenting cells from partially histoincompatible donors. AChR-induced proliferative responses of the purified T cells were restricted to HLA-DR in all six patients (Fig. 3).

In all six patients, the AChR-specific T cells were examined repeatedly for their surface differentiation markers¹⁵. After more than 4 weeks of continuous culture, the cells uniformly (>99%) had the phenotype of activated, proliferating inducer/helper T lymphocytes ($\text{OKT3}^+ \text{OKT4}^+ \text{OKT8}^- \text{OKT11}^+$), as assessed by fluorescence microscopy using mouse monoclonal antibodies against human leukocyte differentiation antigens (Ortho) and standard indirect immunofluorescence techniques. A high proportion (85%) of the cells reacted strongly with several monoclonal antibodies against monomorphic HLA-DR determinants (OKIa1, Ortho; L243, Becton-Dickinson; MAS 053 and 054, Sera-Lab) and also (80–85%) with two monoclonal antibodies against DR β -chains (DA6. 164, DA6. 231; provided by Dr K. Guy¹⁶) and a monoclonal anti-DR3 antibody (16.23; provided by Dr J. Johnson¹⁷).

Functional assays of anti-DR antibodies (dialysed extensively against medium before use in tissue culture) were performed with the DR3-positive long-term cell line. DR3-specific alloantiserum not only stained the line cells by immunofluorescence but also inhibited AChR-induced proliferation (Fig. 4). The monoclonal anti-DR antibodies L243 and MAS 054 were also inhibitory (50% and 80% inhibition, respectively, at 1:125 dilution). Inhibition could still be observed after preincubation

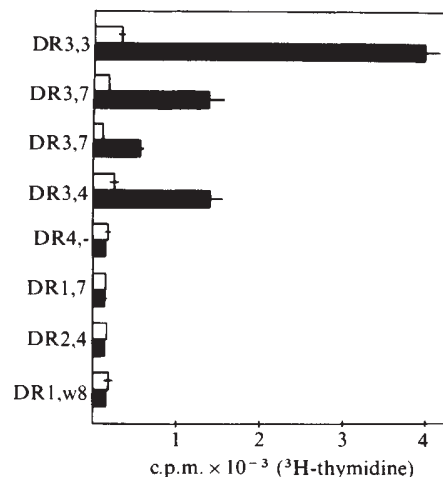


Fig. 3 Genetic restriction by HLA-DR3 and loss of alloreactivity of the autoreactive line T cells (patient 1). 1×10^4 line cells were cultured alone or with 2×10^5 mitomycin C-treated PBM from HLA-typed (standard microcytotoxicity assay) donors as described (Fig. 1) except that the culture medium contained 10% FCS (Gibco) instead of human serum. In the absence of antigen (open columns), background proliferation was low irrespective of the HLA-type of the accessory cells, that is, alloreactivity of the line cells was not detectable. AChR-induced proliferation (solid columns) was absolutely dependent on the presence of accessory cells from donors who shared the HLA-DR3 histocompatibility antigen with the donor of the line cells (DR3, 3; top). Patient 1 is HLA-homozygous as shown by analysis of her parents' HLA phenotype (HLA A1, 2; B8, w44; Cw5, w7; DR3, 4 and HLA A1,-; B8, w44; Cw5, w7; DR1, 3). The five additional patients were typed as DR3, 7 (patient 2), DR2, w8 (patient 3), DR3, 5 (patient 4), DR5,- (patient 5), and DR1, 4 (patient 6). AChR-reactive T cells from these patients were restricted to both parental haplotypes; patient 2 showed a preferential restriction to DR3.

(2 h) of antigen presenting cells with the anti-DR antisera, but not of the responding T cells, indicating that it occurred at the level of the antigen-presenting cells. Irrelevant control antibodies (anti-CALLA, Becton-Dickinson) and monoclonal antibodies against HLA-DC-related determinants (Leu-10, Becton-Dickinson) and against polymorphic DR determinants (MAS 044, Sera-Lab; anti-DR4/DR5/DR7 alloantisera, Biotest, Frankfurt) did not inhibit AChR-induced proliferation.

Functional activity of AChR-reactive T cells was tested in co-cultures with autologous PBM following the protocol outlined in ref. 18. Briefly, PBM or isolated B lymphocytes were cultured in the presence of pokeweed mitogen (PWM; Gibco) or various concentrations of *Torpedo* AChR (10 ng ml^{-1} – $10 \mu\text{g ml}^{-1}$) with or without various numbers of added AChR-specific autologous T cells. Anti-AChR antibody was measured in a modified standard radioprecipitation assay^{19,20} in culture supernatants obtained after 4, 8 and 12 days of culture. So far, only in one of the six patients (patient 1) were anti-AChR autoantibodies detectable in PWM-stimulated cultures (4 fmol ml^{-1} at day 8; controls below 2 fmol ml^{-1}). These findings confirm that AChR-specific B cells may be a limiting factor in the induction of an *in vitro* antibody response by PBM in myasthenia gravis patients²¹. It cannot, therefore, be concluded with certainty that AChR-reactive ('delayed-type hypersensitivity') T cells are functional helper cells. However, their $\text{OKT4}^+ \text{OKT8}^-$ phenotype and their restriction by class II major histocompatibility complex molecules indicates that we have selected regulatory T cells. This would be supported by recent experiments showing that our long-term cell lines have the capacity to produce interleukin-2 (B. Fleischer, unpublished observations). Moreover, previous findings^{22–24} in the experimental autoimmune myasthenia gravis model showed that AChR-reactive T cells exert helper function *in vivo*.

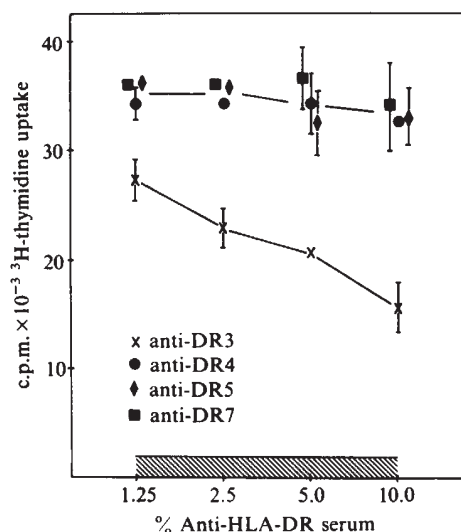


Fig. 4 Inhibition of AChR-induced proliferation (c.p.m. $\pm$ s.d.) of the autoreactive, DR3-positive long-term T-cell line (patient 1) with anti-DR3 alloantisera. The cells were cultured as described in Fig. 1 in the presence of an optimal stimulating concentration of *Torpedo* AChR ($5 \mu\text{g ml}^{-1}$) and various concentrations of thrombocyte-absorbed DR-specific alloantisera (Biotest; anti-DR3: lot no. 111952; anti-DR4: 112042; anti-DR5: 112331; anti-DR7: 112052; all antisera were dialysed against medium before use in tissue culture). Only anti-DR3 antiserum inhibited AChR-induced proliferation. Hatched area: range of background proliferation in the absence of antigen.

Our results suggest that autoreactive inducer T cells or their precursors may initiate and/or regulate the production of myasthenogenic anti-AChR autoantibodies *in vivo*. The finding that the cells react both with *Torpedo* and human AChR suggests that antigenic determinants shared between fish and mammalian AChR at least participate in the primary autoimmunization event^{18,24}, although our propagation protocol (using *Torpedo* AChR) was likely to select for T-cell clones cross-reactive with *Torpedo* and human AChR. The observation that the AChR-specific T cells were, in three patients, restricted either partially or exclusively to HLA-DR3 leads us to speculate that the (weak) association of certain forms of myasthenia gravis with HLA-B8/DR3 (refs 8, 9) may partially reflect the recognition by autoreactive T lymphocytes of self-AChR and, possibly, other autoantigens²⁵ in the context of DR3-related determinants. Purified autoimmune human T cells seem to represent an ideal tool for further studies on autoimmune disease.

We thank our patients for their continuing cooperation; U. Brocke, C. Edler and M. Hendricks for technical assistance; Professor V. P. Whittaker for his gift of *Torpedo* electric organs; Dr B. Fleischer for assaying interleukin-2, Drs K. Guy and J. Johnson for their gift of monoclonal antibodies, Professors H. Kolb and H. Wekerle for their valuable suggestions, and B. Kosmetzky and S. Brinkmann for assistance with the manuscript. This study was supported by the Deutsche Forschungsgemeinschaft, SFB 200, B5.

Received 21 October 1983; accepted 16 April 1984.

1. Lennon, V. A. *Hum. Path.* **9**, 541-551 (1978).
2. Lindstrom, J. M. *Adv. Immun.* **27**, 1-50 (1979).
3. Vincent, A. *Physiol. Rev.* **60**, 756-823 (1980).
4. Drachman, D. B. *A. Rev. Neurosci.* **4**, 195-225 (1981).
5. Toyka, K. V., Drachman, D. B., Pestronk, A. & Kao, I. *Science* **190**, 397-399 (1975).
6. Lindstrom, J. M., Engel, A. G., Seybold, M. E., Lennon, V. A. & Lambert, E. H. *J. exp. Med.* **144**, 739-753 (1976).
7. Castleman, B. *Ann. N.Y. Acad. Sci.* **135**, 496-503 (1966).
8. Compston, D. A. S., Vincent, A., Newsom-Davis, J. & Batchelor, J. R. *Brain* **103**, 579-601 (1980).
9. Grosse-Wilde, H. *et al. Dt. med. Wschr.* **108**, 694-700 (1983).
10. Abramsky, O., Aharonov, A., Webb, C. & Fuchs, S. *Clin. exp. Immun.* **19**, 11-16 (1975).
11. Richman, D., Patrick, J. & Arnason, B. G. W. *New Engl. J. Med.* **294**, 694-698 (1976).
12. Conti-Tronconi, B., Morgutti, M., Sghirlanzoni, A. & Clementi, F. *Neurology* **29**, 496-501 (1979).
13. McQuillen, D. P., Koethe, S. M. & McQuillen, M. P. *J. Neuroimmun.* **5**, 59-65 (1983).

14. Paul, W. E., Sredni, B. & Schwartz, R. H. *Nature* **294**, 697-699 (1981).
15. Reinherz, E. L. & Schlossman, S. F. *New Engl. J. Med.* **303**, 370-373 (1980).
16. Guy, K., Van Heyningen, V., Cohen, B. B., Deane, D. L. & Steel, C. M. *Eur. J. Immun.* **12**, 942-948 (1982).
17. Johnson, J. P., Meo, T., Riethmuller, G., Schendel, D. J. & Wank, R. *J. exp. Med.* **156**, 104-111 (1982).
18. Newsom-Davis, J., Willcox, N. & Calder, L. *New Engl. J. Med.* **305**, 1313-1318 (1981).
19. Lindstrom, J. M., Seybold, M. E., Lennon, V. A., Whittingham, S. & Duane, D. D. *Neurology* **26**, 1054-1059 (1976).
20. Toyka, K. V. *et al. Klin. Wschr.* **57**, 937-942 (1979).
21. Clarke, C., Vincent, A. & Newsom-Davis, J. *J. clin. Sci.* **56**, 1 Abstr. (1979).
22. Hohlfeld, R., Kalies, I., Heinz, F., Kalden, J. R. & Wekerle, H. *J. Immun.* **126**, 1355-1359 (1981).
23. Hohlfeld, R., Kalies, I., Ernst, M., Ketelsen, U. P. & Wekerle, H. *J. neurol. Sci.* **57**, 265-280 (1982).
24. Wekerle, H., Hohlfeld, R., Ketelsen, U. P., Kalden, J. R. & Kalies, I. *Ann. N.Y. Acad. Sci.* **377**, 455-476 (1981).
25. Ryder, L. P., Anderson, E. & Svegaard, A. *HLA and Disease Registry* 3rd edn (Munksgaard, Copenhagen, 1979).
26. Kalies, I., Kalden, J. R., Heinz, F., Janzen, R. W. C. & Lachenmayer, L. *Klin. Wschr.* **57**, 875-881 (1979).
27. Kurnick, J. T. *et al. Scand. J. Immun.* **10**, 563-573 (1979).

Development of X- and Y-cell retinogeniculate terminations in kittens

Mriganka Sur*, Rosalyn E. Weller & S. Murray Sherman†

Department of Neurobiology and Behavior, State University of New York at Stony Brook, Stony Brook, New York 11794, USA

The cat retinogeniculocortical pathways are organized chiefly into two parallel independent neuronal streams, one involving X-cells of the retina and lateral geniculate nucleus, and the other, Y-cells^{1,2}. Development of the Y-cell pathway is more seriously affected by visual deprivation than is the X-cell pathway^{2,3} and we reasoned that some insight into the underlying mechanisms of these effects could be gained from studies of normal development. We therefore injected horseradish peroxidase into physiologically identified X- and Y-cell retinogeniculate axons to examine the postnatal development of their terminations in kittens. As we report here, at 3-4 weeks of age, most optic tract axons can be identified physiologically as members of the X- or Y-cell class. X-cell terminal fields in lamina A or AI are wider at 3-4 weeks than they are in adults, while Y-cell terminal fields are narrower than in adults^{4,5}. During the second and third postnatal months, X-cell terminal arbors progressively contract while Y-cell arbors expand so that, by 12 weeks of age, the adult pattern is seen. These data, and the results of our earlier study of the effects of monocular lid suture on these terminal arbors³, suggest that enlargement of Y-cell terminations in geniculate lamina A or AI during development may be accompanied by competitive pruning of X-cell terminations within these same laminae.

Experiments were performed on three groups of kittens: 17 at 21-30 days old (the 3-4-week old group), 5 at 51-62 days old (the 8-week old group) and 5 at 81-91 days old (the 12-week old group). Animals in the 3-4- and 8-week old groups were anaesthetized with 1-2% halothane and paralysed with gallamine triethiodide; a respirator was used to hyperventilate them slightly with a mixture of 70% nitrous oxide, 29% oxygen and 1% carbon dioxide^{6,7}. Our physiological and morphological methods for the 12-week old kittens were the same as those described previously for adult cats^{4,8}. Briefly, micropipettes containing horseradish peroxidase (HRP) were used to record retinal fibres in the lateral geniculate nucleus (LGN) or in the subjacent optic tract. Each axon was identified as an X- or Y-cell on a battery of tests, including receptive field size, linearity of spatial and temporal summation, and axonal conduction velocity^{9,10} (the application of these tests to kittens is described more fully below). The axon was then impaled and injected with HRP. A total of 190 retinal axons were recorded in these kittens and

* Present address: Section of Neuroanatomy, Yale University School of Medicine, New Haven, Connecticut 06510, USA.

† To whom correspondence should be addressed.

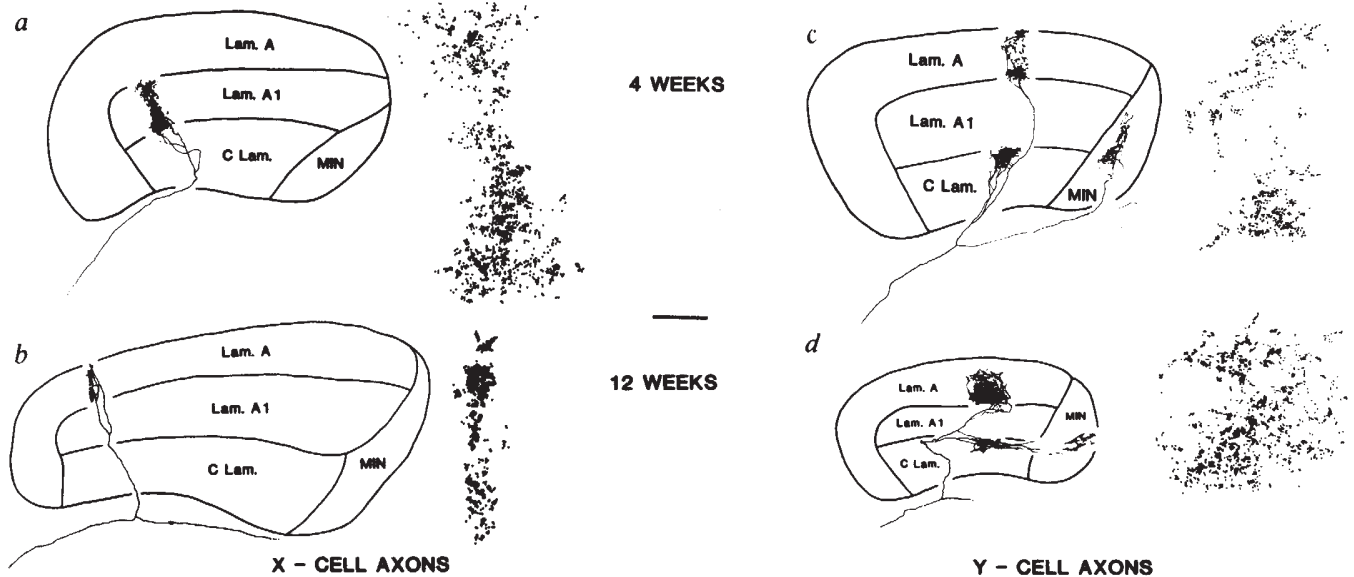


Fig. 1 Retinogeniculate arbors from single axons in kittens. Each axon is illustrated by a pair of drawings: left, a lower-power drawing of the arbor in the LGN; right, a higher-power drawing limited to the terminal boutons in lamina A or A1. The scale bar in the centre of the figure applies to all pairs (a–d) and represents 500 μm for the lower-power drawings and 100 μm for the higher-power ones. *a*, X-cell axon from the ipsilateral retina in a 30-day old kitten. The cell has an on-centre receptive field 2.0° in diameter located 26° from the vertical meridian and 38° below the horizontal zero parallel. Its latency to optic chiasm stimulation was 1.6 ms. Its responses to a counterphased, sine-wave grating exhibited linear spatial and temporal summation. This axon had 1,073 boutons in lamina A1. For comparison, X-cell axons from normal adult cats similarly analysed had 526–915 boutons in lamina A or A1⁴. *b*, Y-cell axon from the contralateral retina in a 23-day old kitten. The cell had an on-centre receptive field 3.3° in diameter located 5° from the vertical meridian and 6° below the horizontal zero parallel. Its latency to optic chiasm stimulation was 1.1 ms and its responses to a counterphased, sine-wave grating exhibited nonlinear summation. The axon had 565 boutons in lamina A, whereas 4 Y-cell axons from normal, adult cats similarly analysed had 651–1,405 boutons in lamina A or A1⁴. *c*, X-cell axon from the contralateral retina in an 81-day old kitten. The cell had an on-centre receptive field 0.9° in diameter located 35° from the vertical meridian and 1° below the horizontal zero parallel. Its latency to optic chiasm stimulation was 1.2 ms and it responded linearly to gratings. The terminal field in lamina A had 573 boutons. *d*, Y-cell axon from the contralateral retina in an 86-day old kitten. The cell had an off-centre receptive field 2.0° in diameter located 5° from the vertical meridian and 21° below the horizontal zero parallel. Its latency to optic chiasm stimulation was 0.6 ms and it responded nonlinearly to gratings. The terminal field in lamina A had 1,170 boutons.

of these, 65 were successfully filled with HRP and morphologically analysed. Brains were sectioned coronally or sagittally at 100 μm and reacted with 3-3-diaminobenzidine with cobalt intensification¹¹.

Physiologically, the 75 retinal axons recorded in 3–4-week old kittens could follow high frequencies (>100 Hz) of optic chiasm stimulation but had longer latencies to such stimulation than did axons in older kittens or adults. Although nonlinear responses to counterphased sine-wave gratings were sometimes variable in young kittens^{7,12,13}, these responses could nonetheless be reliably detected when 'null positions'¹⁰ were used to minimize the linear response component. In the 3–4-week old kittens, the overwhelming majority of axons (46 of 49) identified as Y-cells had frequency-doubled, nonlinear response. Twenty-six axons identified as X-cells exhibited only linear responses. Y-cells exhibited a shorter latency to optic chiasm stimulation (mean 0.9 ms, range 0.6–1.2 ms) than did X-cells (mean 1.7 ms, range 1.2–2.5 ms). Receptive field centres for these cells in 3–4-week old kittens were roughly four times as large as in adults (mean for kitten X-cells 2.1° , range 1.0 – 5.0° ; mean for kitten Y-cells 4.0° , range 1.7 – 7.5° ; see ref. 12) and lacked detectable antagonistic surrounds. Y-cell receptive field centres were larger than X-cell centres at similar eccentricities. The three axons exhibiting only linear responses which were identified as Y-cells had chiasm latencies (0.9–1.1 ms) and field centres (3 – 4.5°) comparable with nonlinear Y-cells. Such 'linear' Y-cells have also been described in the kitten LGN¹⁰. None of these three was recovered histologically.

By 8 weeks of age, receptive field properties of the 54 recorded axons (29 X-cells and 25 Y-cells) were more adult-like than at 3–4 weeks of age. Compared with receptive fields of the younger kittens, those of 8-week old kittens were smaller and often comparable with adult sizes (see ref. 14), their surrounds were more developed and the nonlinear responses of the Y-cells were

stronger and more stable. At 12 weeks of age, the 30 X-cells and 31 Y-cells were physiologically almost indistinguishable from those in adults^{10,15}.

Morphologically, X- and Y-cell terminations in 3–4-week old kittens clearly differed from those in adults both in the geometry of terminal zones in the A-laminae and in the sizes and shapes of individual terminal boutons. These differences were less pronounced at 8 weeks of age and by 12 weeks of age terminal fields appeared quite adult-like.

We successfully injected and recovered 10 X-cell and 12 Y-cell retinogeniculate axons in 3–4-week old kittens. Figure 1*a,c* illustrates representative X- and Y-cell terminal fields from these kittens. The X-cell axon arises from the ipsilateral retina, its terminal field is confined to lamina A1 of the LGN and it has a large number of boutons within a broad terminal field. The Y-cell axon from the contralateral retina innervates laminae C and A as well as the medial interlaminar nucleus of the LGN. Its terminal field in lamina A is quite narrow and sparse, with fewer boutons, compared with either X-cell arbors at similar ages or Y-cell terminal arbors in adults.

We have also recovered 12 X- and 7 Y-cell retinogeniculate axons from 8-week old kittens and 12 X- and 12 Y-cell axons from 12-week old kittens. Figure 1*b,d* illustrates representative arbors of an X- and a Y-cell from a 12-week old kitten. These arbors are not obviously different from those in adults⁴, but they do differ from arbors in younger kittens. The X-cell terminal arbor in the A-laminae is much broader in the 3–4-week old kitten than in the 12-week old kitten, while the Y-cell arbor in the A-laminae is narrower and sparser in the 3–4-week old kitten than in the 12-week old kitten.

Figure 2 summarizes the postnatal development of terminal arbor widths in the A-laminae of X- and Y-cell retinogeniculate axons and also shows data from 16 X- and 12 Y-cell axons from adult cats⁴. X-cell axons gradually decrease in terminal arbor

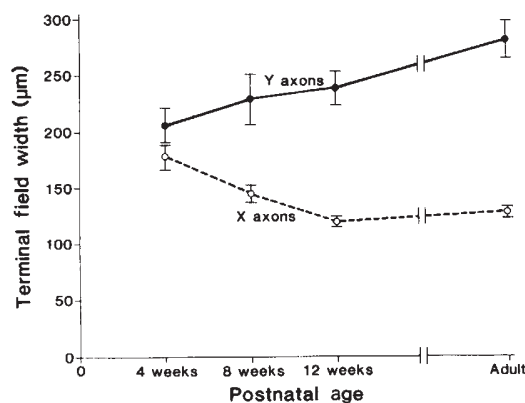


Fig. 2 Maximum width of axon terminal arbors (mean $\pm$ s.d.) in lamina A or AI as a function of postnatal age.

width from 3–4 weeks of age until 12 weeks of age, while Y-cell axons increase in width to 12 weeks of age and beyond to adult values. Both trends are significant ($P < 0.01$ on a χ^2 test). Furthermore, the X- and Y-cell widths at 3–4 weeks are each significantly different from the corresponding values in 12-week old kittens or adults ($P < 0.05$ for the X-cell comparison between 3–4 and 12 weeks of age; $P < 0.01$ for the other comparison; Mann-Whitney U -test).

In addition to these postnatal changes in the geometry of terminal arbors, individual boutons also mature between 3 and 12 weeks postnatally^{16,17}. Figure 3a illustrates a portion of the terminal field of an X-cell axon from a 27-day old kitten. Terminals at this age are often small and irregularly shaped compared with the well formed terminals seen at 12 weeks of age (Fig. 3b). Many terminals in the 3–4-week old kittens have spike-shaped or filopodial endings with growth-cone-like extensions (see ref. 17). Nonetheless, the clustering of terminals characteristic of X-cell arbors in 12-week old kittens (Fig. 3b) or in adult cats⁴ already exists in the younger kittens studied (Fig. 3a). A Y-cell arbor in a 23-day old kitten (Fig. 3c, d) exhibits terminal morphology similar to that of the immature X-cell axon. In particular, Y-cell axons also exhibit numerous filopodial endings and terminations, and the swellings themselves appear generally smaller than those in 12-week old kittens (Fig. 3e) or adults⁴.

The present results indicate that, during the critical period for visual development¹⁸, there is dynamic shaping of the terminal boutons and terminal arbors of retinal X- and Y-cell axons in the LGN. Compared with retinal Y-cells, retinal X-cells seem to be generated earlier¹⁹ and may innervate the LGN earlier^{6,20}. Therefore, by a process analogous to that suggested for other developing systems²¹, X-cell axons may initially have abundant terminations that retract or become pruned as Y-cells subsequently establish their arbors in the LGN, and the retraction of X-cell arbors and expansion of Y-cell arbors may occur by a process of mutual competition. Because much of the Y-cell expansion occurs during the critical period, eyelid suture during this time may interfere with the ability of Y-cell arbors to compete successfully for terminal space with the already developed X-cell arbors. Indeed, within the A-laminae, many X- and Y-cell arbors from the deprived eye of monocularly sutured cats exhibit geometry similar to that in 3–4-week old kittens and unlike that in adults³. It is also possible that the normal growth of Y-cell arbors and shrinkage of X-cell arbors are developmental processes independent of one another, that they involve no competitive interactions and that deprivation simply retards these developmental processes. However, Sur *et al.*³ showed that retinogeniculate Y-cell axons from the deprived eye develop normally where X-cell arbors are never substantively present (that is, in lamina C) and that these Y-cell arbors fail to develop normally only in the presence of X-cell arbors (that is, in the A-laminae). This suggests that the development of

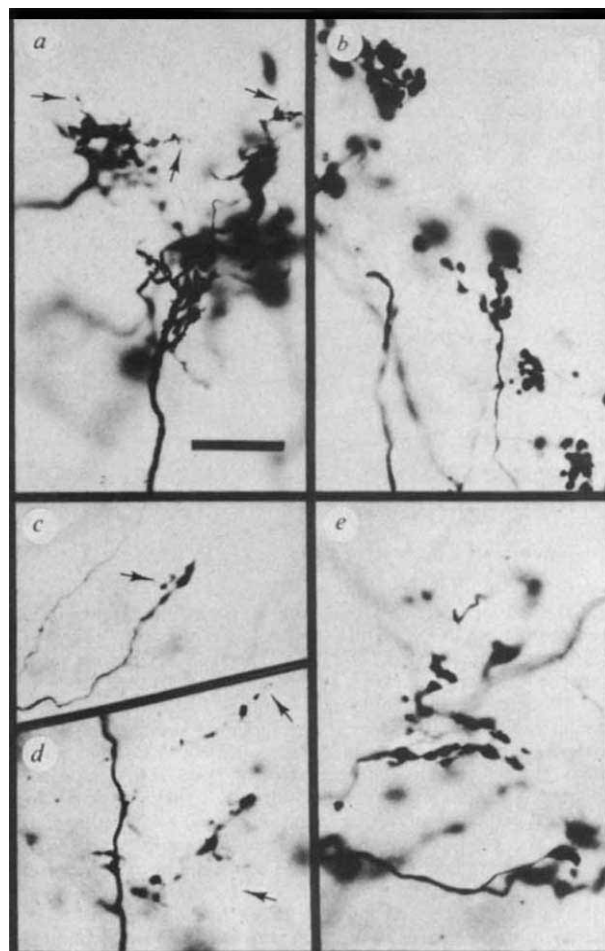


Fig. 3 Morphology of retinogeniculate terminal endings in kittens. a, Portion of X-cell terminal field in a 27-day old kitten. Arrows indicate filopodial or spike-shaped endings characteristic of terminals at this age. Scale bar, 25 μ m, applying to b–e also. b, Portion of X-cell terminal field in a 91-day old kitten. Terminals have the form and features characteristic of adult X-cell axons. c, d, Portions of Y-cell terminal field in a 23-day old kitten. Arrows again indicate filopodial endings. Y-cell terminal fields are much sparser than X-cell arbors at this age. e, Portion of Y-cell terminal arbor in a 91-day old kitten showing adult-like terminal morphology.

retinogeniculate X- and Y-cell arbors is truly a competitive process.

Finally, we note that the pruning of X-cell arbors and growth of Y-cell arbors within individual lamina A or AI probably occurs independently of the segregation of afferents from the two eyes within the LGN. The latter occurs prenatally²² and does not seem to involve significant initial proliferation and later pruning of terminal fields²³. Interactions between X- and Y-cell retinogeniculate terminations from the same eye appear to be superimposed postnatally on an afferent population that has already segregated into geniculate laminae.

We thank Joan Sommermeyer and Angelo Gero for technical assistance. This work was supported by USPHS grants EY05241 (M.S.), EY05674 (R.E.W.) and EY03038 (S.M.S.).

Received 24 January; accepted 10 May 1984.

1. Stone, J., Dreher, B. & Leventhal, A. *Brain Res. Rev.* **1**, 345–394 (1979).
2. Sherman, S. M. & Spear, P. D. *Physiol. Rev.* **62**, 738–855 (1982).
3. Sur, M., Humphrey, A. L. & Sherman, S. M. *Nature* **300**, 183–185 (1982).
4. Sur, M., & Sherman, S. M. *Science* **218**, 389–391 (1982).
5. Bowling, C. B., & Michael, D. R. *Nature* **286**, 899–902 (1980).
6. Daniels, J. D., Pettigrew, J. D. & Norman, J. L. *J. Neurophysiol.* **41**, 1373–1393 (1978).
7. Wilson, J. R., Tessin, D. E. & Sherman, S. M. *J. Neurosci.* **2**, 562–571 (1982).
8. Friedlander, M. J., Lin, C.-S., Stanford, L. R. & Sherman, S. M. *J. Neurophysiol.* **46**, 80–129 (1981).
9. Hoffmann, K.-P., Stone, J. & Sherman, S. M. *J. Neurophysiol.* **35**, 518–531 (1972).
10. Hochstein, S. & Shapley, R. M. *J. Physiol., Lond.* **262**, 237–264 (1976).

11. Adams, J. C. *Neuroscience* **2**, 141–145 (1977).
12. Mangel, S. C., Wilson, J. R. & Sherman, S. M. *J. Neurophysiol.* **50**, 240–262 (1983).
13. Hamasaki, D. I. & Sutija, V. G. *Expl Brain Res.* **35**, 9–23 (1979).
14. Cleland, B. G. & Levick, W. R. *J. Physiol., Lond.* **240**, 421–456 (1974).
15. Rusoff, A. C. & Dubin, M. W. *J. Neurophysiol.* **40**, 1188–1198 (1977).
16. Friedlander, M. J., Vahle-Hinz, C. & Martin, K. A. C. *Invest. ophthalm. vis. Sci. (ARVO Abstr.)* **24**, Suppl. 138 (1983).
17. Mason, C. A. *Neuroscience* **7**, 541–559 (1982).
18. Hubel, D. H., & Wiesel, T. N. *J. Physiol., Lond.* **206**, 419–436 (1970).
19. Walsh, C., Polley, E. H. & Hickey, T. L. *Nature* **302**, 611–614 (1983).
20. Torrealba, F., Guillery, R. W., Eysel, U., Polley, E. H. & Mason, C. A. *J. comp. Neurol.* **211**, 377–396 (1982).
21. Purves, D. & Lichtman, J. R. *Science* **210**, 153–157 (1980).
22. Shatz, C. J. *J. Neurosci.* **3**, 482–499 (1983).
23. Sretavan, D. W. & Shatz, C. J. *Nature* **308**, 845–848 (1984).

Expression of *myb*, *myc* and *fos* proto-oncogenes during the differentiation of a murine myeloid leukaemia

Thomas J. Gonda* & Donald Metcalf†

* Melbourne Tumour Biology Branch, Ludwig Institute for Cancer Research and † The Walter and Eliza Hall Institute of Medical Research, PO Royal Melbourne Hospital, Victoria 3050, Australia

It is widely thought that *c-onc* genes (or proto-oncogenes)—the cellular progenitors of retroviral transforming genes—are involved in cellular differentiation and/or proliferation. Such ideas originate primarily from the ability of *v-onc* genes and 'activated' *c-onc* genes to induce uncontrolled cellular proliferation, and their capacity to arrest or interfere with differentiation processes in some systems. Haematopoietic cell populations provide additional support for these ideas as *c-myb* RNA is present in cell lines corresponding to immature, but not mature, cell types^{1,2}, and elevated levels have been found in tissues that are active in haematopoiesis^{2,3}. We have now examined the effects of induced differentiation on *c-onc* gene expression in a murine myeloid leukaemia cell line, WEHI-3B (D⁺ subline)⁴. Our results show that the expression of *c-myb* and *c-myc*, at the level of transcription, decreases only at late stages in the monocytic differentiation of WEHI-3B cells, while expression of *c-fos* increases markedly. We suggest that *c-myb* and *c-myc* do not themselves control myeloid differentiation, but that they function in the maintenance of the proliferative state of myeloid cells. The induction of *c-fos* may reflect its role in some macrophage-specific functions.

The cloned myelomonocytic leukaemia WEHI-3B can be induced to differentiate to both monocytes (macrophages) and granulocytes by granulocyte-colony stimulating factor⁴ (G-CSF^{5,6}). However, because G-CSF alone does not induce complete differentiation in cultures of a few days' duration, we have used G-CSF plus a low concentration of actinomycin D, a combination which Cooper *et al.*⁷ have shown efficiently induces predominantly monocytic differentiation of WEHI-3B cells. In our hands, G-CSF plus actinomycin D routinely induced differentiation in ~90% of the WEHI-3B cells after 2 days' exposure. However, the degree of differentiation at this time varied between experiments, with differing proportions of promonocytes (Fig. 1b) and mature monocytes (Fig. 1c) being present (Table 1, expts 2–4). In all experiments, most cells were classified as mature monocytes by 3 to 4 days.

In order to examine *c-onc* expression in WEHI-3B cells, polyadenylated RNA was fractionated by formaldehyde-agarose gel electrophoresis and transferred to nitrocellulose filters. Hybridization to radiolabelled probes derived from cloned *onc* genes revealed the presence of transcripts corresponding to *c-abl*, *c-Ki-ras*, *c-fms*, *c-fes*, *c-Ha-ras*, *c-myb*, *c-myc* and *c-fos*. The *c-abl* and *c-Ki-ras* transcripts (Fig. 2) were of similar sizes to those reported elsewhere (refs 8 and 9, respectively). WEHI-3B cells contain two *c-fms* transcripts of 4.1 and 8.4 kilobases (kb) (Fig. 2), the smaller of which probably corresponds to the 3.7-kb transcript seen in mouse placental tissue¹⁰. The *c-fes* gene is also transcribed in WEHI-3B, giving rise to a single 3.2-kb RNA species. Although the size of the *c-fes* tran-

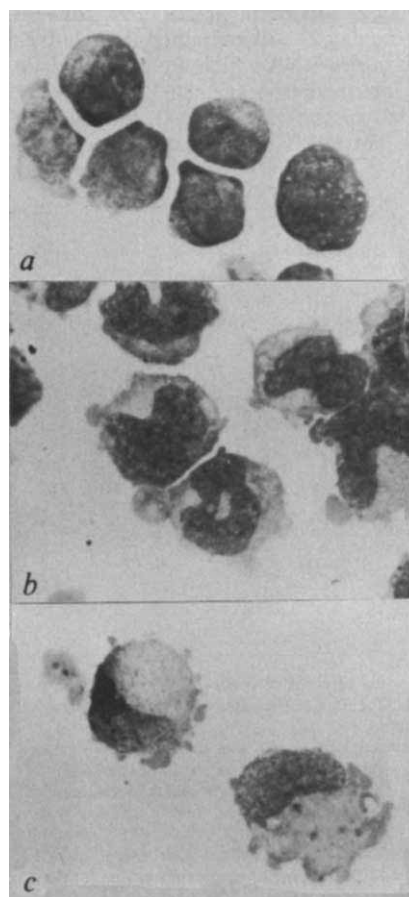


Fig. 1 Morphology of WEHI-3B cells undergoing monocytic differentiation. *a*, Undifferentiated WEHI-3B cells; *b*, promonocytes; *c*, monocytes appearing after induction of differentiation. **Methods:** WEHI-3B cells (D⁺ subline)⁴ were grown as described previously and treated with 370 U ml⁻¹ G-CSF (partially purified by gel filtration from post-endotoxin serum)^{5,6} plus 5 ng ml⁻¹ actinomycin D. Cyto-centrifuge preparations were stained with May-Grünwald/Giemsa.

script has not previously been reported, we note that the avian homologue of *c-fes* (*c-fps*) is transcribed¹¹ and its product expressed¹² in haematopoietic cells. Transcription of *c-fes* and *c-fms* is under investigation and will not be dealt with further here. No transcripts of *c-src*, *c-sis*, *c-mos*, *c-erb-A* or *c-erb* were detected.

Marked changes were observed in the levels of *c-myb*, *c-myc* and *c-fos* RNAs following induction of differentiation. For example, expt 1 (Table 1, Fig. 3) shows that differentiation of WEHI-3B cells to monocytes after 3 days of treatment with G-CSF plus actinomycin D resulted in a substantial decrease in the levels of *c-myb* and *c-myc* (6-fold and 10-fold, respectively, determined by optical densitometry of the autoradiograms), while transcription of *c-fos* increased (11-fold) and *c-Ha-ras* transcription was essentially unaltered.

To further correlate these changes in *c-onc* gene expression with different stages of WEHI-3B cell maturation, cells were collected after varying periods of induction, and independent experiments were compared in which the time course of differentiation had varied (Table 1, Fig. 3, expts 2–4). In expt 2, in which induction of differentiation for 2 days resulted in the differentiation of most of the cells to promonocytes, no significant changes in *c-myb* or *c-myc* RNA levels were observed. Furthermore, in expt 3, where most of the cells were classified as mature monocytes after 2 days' treatment (confirmed by staining cells for nonspecific esterase activity¹³ and by monitoring the suppression of clonogenicity in soft agar cultures⁴), only a slight decrease in the levels of *c-myb* and *c-myc* was detected. This experiment shows that cells which had only just reached

the monocyte stage at day 2 (few mature monocytes were ever observed before day 2) still contained the relatively high levels of *c-myb* and *c-myc* RNA which are characteristic of undifferentiated and intermediate-stage cells. However, by 3 to 4 days' treatment in all cases (expts 1, 3, 4), the levels of *c-myb* and *c-myc* RNAs had fallen dramatically (Fig. 3; *c-myb* data for

expt 4 not shown), implying that transcription of *c-myb* and *c-myc* had ceased during the 24–48 h following the appearance of mature monocytes. Thus, taken together, these results imply that the decrease in *c-myb* and *c-myc* transcription occurs only after the differentiating cells reach the monocyte stage.

In view of the decreased or unaltered expression of several of the *c-onc* genes during WEHI-3B differentiation, we were surprised by the dramatic increase in the level of *c-fos* transcription (Table 1, Fig. 3, expts 1, 4). The expression of *c-fos* is probably associated with mature monocytes (see also below), as the rise in *c-fos* transcription paralleled the appearance of mature cells in expt 4, in which only a small increase in the total of all differentiated cells occurred between days 2 and 4.

The rationale of this study was that changes in *c-onc* expression which occur during WEHI-3B differentiation would reflect changes that occur in normal myeloid differentiation. Thus, we would predict that normal granulocyte-macrophage progenitor cells (GM-CFCs)¹⁴ would express similar levels of *c-myb* and *c-myc* RNAs to WEHI-3B and that normal macrophages would express low levels of *c-myb* and *c-myc* and high levels of *c-fos* RNA. The first prediction could not be readily tested due to the difficulty in obtaining sufficient numbers of pure GM-CFCs; however, a second prediction, that other early myeloid cell lines and populations containing myeloid progenitors or stem cells should express *c-myb* and *c-myc*, could be tested. Thus, comparable levels of *c-myc* and *c-myb* RNA were found in another early myeloid cell line, 416B (ref. 15), in a multipotential (myeloid-erythroid-megakaryocytic) cell line B6-SutA C1.15 (ref. 16) and in a mixed population of haematopoietic cells from fetal liver¹⁴ (a low-density fraction containing about 5% GM-CFCs) (Fig. 4). Normal macrophages, however, contained no detectable *c-myb* RNA and a low level of *c-myc* RNA (Fig. 4). By contrast, a high level of *c-fos* RNA was found in the macrophages and also in the fetal liver cells, but not in the immature cell lines (Fig. 4).

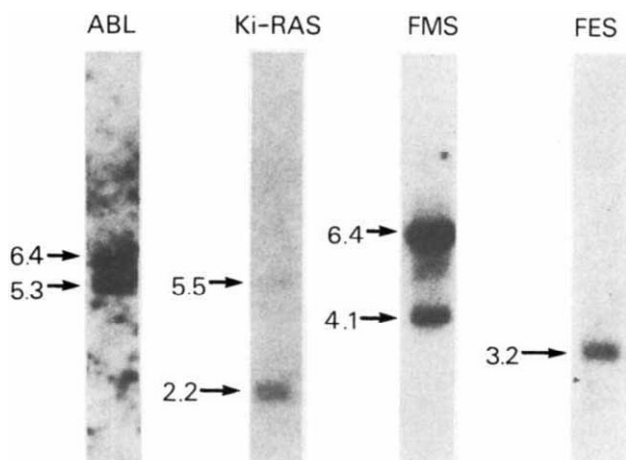


Fig. 2 Transcripts of *c-abl*, *c-Ki-ras*, *c-fms* and *c-fes* in undifferentiated WEHI-3B cells. The sizes, in kilobases, of the major transcripts for each *c-onc* gene are indicated.

Methods: Polyadenylated RNA was prepared and analysed by formaldehyde-agarose gel electrophoresis as previously described². RNA was then transferred on to nitrocellulose filters which were then hybridized with ³²P-labelled nick-translated²⁴ probes, washed and exposed to X-ray film. Hybridizations were performed for 2–3 days at 41 °C in a buffer containing 50% formamide and 3 × SSC and filters were washed at 50 °C in 0.1 × SSC, 0.1% SDS. Cloned DNAs used to prepare the probes for *c-abl*, *c-Ki-ras* and *c-fms* were those used in refs 8, 9 and 10, respectively; the *c-fes* probe was the 0.7-kb *Pst*I fragment 'S₁' of Snyder-Theilen feline sarcoma virus²⁵.

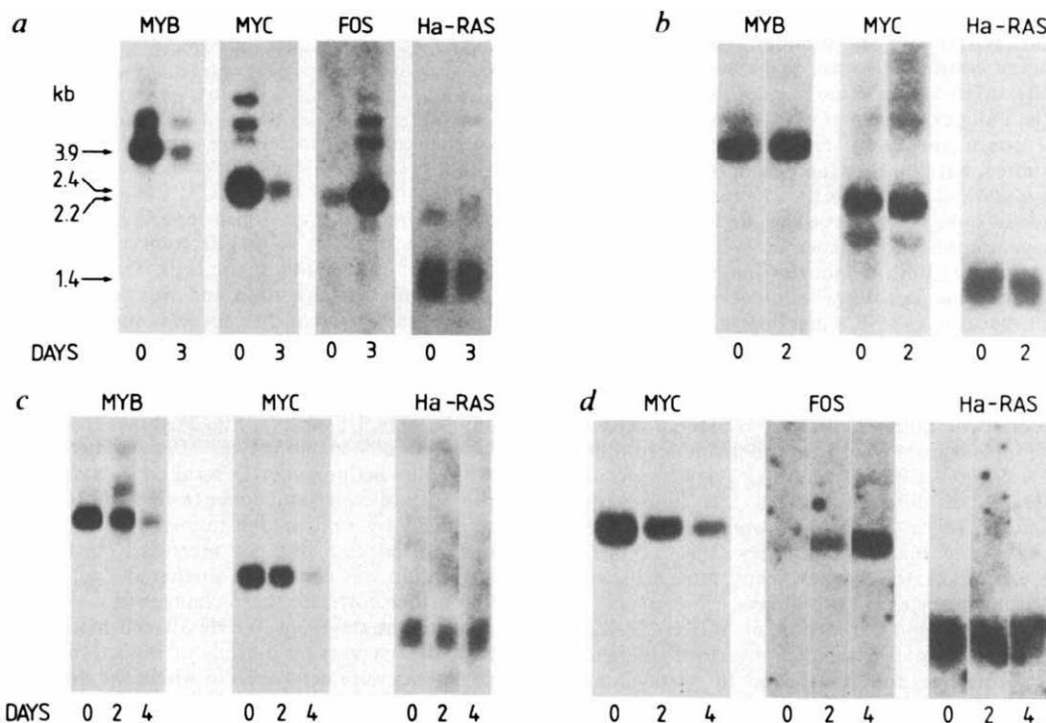


Fig. 3 Transcripts of *c-myb*, *c-myc*, *c-fos* and *c-Ha-ras* in undifferentiated and differentiated WEHI-3B cells. Data are from the experiments 1–4 detailed in Table 1. Days refer to the time of treatment with inducers. The sizes of the major transcripts for each *c-onc* gene are shown at the left of panel 1. The larger transcripts are probably nuclear precursors; for *c-myb*, a similar-sized precursor has been detected in chicken cells²⁶. The smaller RNA seen with the *myc* probe in expt 2 is not *c-myc*-specific as it is not detected with the homologous murine probe.

Methods: Electrophoresis of polyadenylated RNA, transfer and hybridization were as described in Fig. 2 legend, except that where *v-myb* or *v-myc* probes were used (see below), hybridizations were at 37 °C in buffer containing 5 × SSC and 40% formamide, and filters were washed at 50 °C in 0.5 × SSC, 0.1% SDS. Cloned DNAs used to prepare the probes were *myc*: either the *v-myc*-specific *Pst*I fragment of cloned MC29 proviral DNA²⁷ or the *Bam*HI–*Hind*III fragment containing exons 2 and 3 of the murine *c-myc* gene²⁸; *myb*: either the *Kpn*I–*Xba*I fragment of cloned avian myeloblastosis virus proviral DNA^{26,29} or cloned murine *c-myb* cDNA (N. Gough, A. Dunn and T.J.G., unpublished); *fos*: the 1.0-kb *v-fos*-specific *Pst*I fragment (from pfos-1) of cloned FBJ murine osteosarcoma virus proviral DNA³⁰; and *ha-ras*: a 0.45-kb *v-Ha-ras*-specific fragment of cloned Harvey murine sarcoma virus (clone BS-9)³¹.

What, then, do the results presented here suggest about the possible roles of *c-myb*, *c-myc* and *c-fos* in myeloid differentiation? Clearly, expression of *c-myb* and *c-myc* at high levels does not prevent many of the cellular changes indicative of differentiation, nor do changes in the expression of these genes necessarily occur early in the differentiation process. With regard to the latter point, Reitsma *et al.*¹⁷ have recently reported that the level of *c-myc* RNA in the human myelomonocytic cell line, HL-60, decreases very soon after induction of differentiation with a vitamin D derivative (about twofold after 4 h). In view of our results, we suggest that the rapid 'switch-off' of *c-myc* in HL-60 may be a characteristic of the cell line and/or inducer, rather than a mandatory early event in myeloid differentiation. Our observation that differentiation can proceed in the presence of *c-myb* expression is consistent with work by Durban and Boettiger on avian myeloblastosis virus (AMV), which carries the *v-myb* oncogene. These authors reported that AMV can transform both mature macrophages¹⁸ and macrophage progenitors¹⁹, and furthermore that transformation of progenitor cells occurs only after the cells have differentiated to adherent (mature) macrophages. Thus, it seems that myeloid differentiation can also occur in the presence of *v-myb* expression.

Hence, it seems unlikely that *c-myb* or *c-myc* has a controlling function in the differentiation of myeloid cells. Rather, the expression of these genes in immature cells may reflect a function related to the proliferative state of myeloid (and other haematopoietic) cells. This notion is supported by the recent observation that *c-myc* transcription is stimulated by mitogens in both lymphocytes and fibroblasts²⁰, and indirectly by the finding of a degree of relatedness between the *myb* and *myc* gene products²¹. Thus, the greatly reduced expression of *c-myb* and *c-myc* in mature cells may reflect the diminished proliferative

capacity of these cells compared with undifferentiated and immature cells. In addition, *c-myb* and *c-myc* may also affect expression of other genes, since avian macrophages transformed by *v-myb* and *v-myc* differ in the expression of several differentiation markers²².

To date, *c-fos* transcription has been detected primarily in extraembryonic tissue and in bone and skin tissues²³. The increase in *c-fos* transcription described here, which correlates with monocytic differentiation in both leukaemic and normal cells, suggests that *c-fos* may also be required for the expression of some macrophage-specific functions.

We thank Dr A. W. Burgess for providing G-CSF, Dr J. Hamilton for providing bone marrow-derived macrophages, Drs J. Adams, A. Dunn, N. Gough and K. Savin for valuable comments on the manuscript, Ms D. Grail for technical assistance and Ms S. Blackford for typing the manuscript.

Received 20 March; accepted 21 May 1984.

- Westin, E. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 2194–2198 (1982).
- Gonda, T. J., Sheiness, D. K. & Bishop, J. M. *Molec. cell. Biol.* **2**, 617–624 (1982).
- Chen, J. H. *J. Virol.* **36**, 162–170 (1980).
- Metcalfe, D. & Nicola, N. A. *Int. J. Cancer* **30**, 773–780 (1982).
- Nicola, N. A., Metcalfe, D., Matsumoto, M. & Johnson, G. R. *J. biol. Chem.* **258**, 9017–9023 (1983).
- Burgess, A. W. & Metcalfe, D. *Int. J. Cancer* **26**, 647–654 (1980).
- Cooper, P. C., Metcalfe, D. & Burgess, A. W. *Leuk. Res.* **6**, 313–327 (1982).
- Wang, J. in *Mechanisms of B-Cell Neoplasia*, 157–166 (Editiones Roche, Basel, 1983).
- Ellis, R. W., DeFeo, D., Furth, M. E. & Scolnick, E. M. *Molec. cell. Biol.* **2**, 1339–1345 (1982).
- Müller, R. *et al. Molec. cell. Biol.* **3**, 1062–1069 (1983).
- Shibuya, M., Hanafusa, H. & Balduzzi, P. C. *J. Virol.* **42**, 143–152 (1982).
- Mathey-Prevot, B., Hanafusa, H. & Kawai, S. *Cell* **28**, 897–906 (1982).
- Yam, L. T., Li, C. Y. & Crosby, W. H. *Am. J. clin. Path.* **55**, 283–290 (1971).
- Nicola, N. A., Metcalfe, D., von Melchner, H. & Burgess, A. W. *Blood* **58**, 376–386 (1981).
- Dexter, T. M., Allen, T. D., Scott, D. & Teich, N. M. *Nature* **277**, 471–474 (1979).
- Greenberger, J. S., Sakakeeny, M. A., Humphries, R. K., Eaves, C. J. & Eckner, R. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2931–2935 (1983).
- Reitsma, P. H. *et al. Nature* **306**, 492–494 (1983).
- Durban, E. M. & Boettiger, D. J. *J. Virol.* **37**, 488–492 (1981).
- Boettiger, D. & Durban, E. M. *Cold Spring Harb. Symp. quant. Biol.* **44**, 1249–1254 (1979).
- Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603–610 (1983).
- Ralston, R. & Bishop, J. M. *Nature* **306**, 803–806 (1983).
- Durban, E. M. & Boettiger, D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3600–3604 (1981).
- Müller, R., Slamon, D. J., Tremblay, J. M., Cline, M. J. & Verma, I. M. *Nature* **299**, 640–644 (1982).
- Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **113**, 237–251 (1977).
- Franchini, G., Even, J., Scherr, C. J. & Wong-Staal, F. *Nature* **290**, 154–157 (1981).
- Gonda, T. J. & Bishop, J. M. *J. Virol.* **46**, 212–220 (1983).
- Vennstrom, B., Moscovici, C., Goodman, H. M. & Bishop, J. M. *J. Virol.* **29**, 625–631 (1981).
- Adams, J. M., Gerondakis, S., Webb, E., Corcoran, L. M. & Cory, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1982–1986 (1983).
- Klempner, K.-H., Gonda, T. J. & Bishop, J. M. *Cell* **31**, 453–463 (1982).
- Curran, T., Peters, G., Van Beveren, C., Teich, N. M. & Verma, I. M. *J. Virol.* **44**, 674–682 (1982).
- Ellis, R. W. *et al. J. Virol.* **36**, 408–420 (1980).
- Tushinski, R. J. *et al. Cell* **28**, 71–81 (1982).

Table 1 Differentiation of WEHI-3B cells induced by G-CSF plus actinomycin D

Expt	Days of induction	% Differentiation	
		Promonocytes	Monocytes
1	0	4	0
	3	11	87
2	0	6	1
	2	84	5
3	0	2	1
	2	13	84
	4	13	87
4	0	2	0
	2	70	7
	4	37	60

Results of differential counts for four separate experiments are shown; differentiation was assessed by examination of stained cytocentrifuge preparations. Methods were as described in Fig. 1 legend.

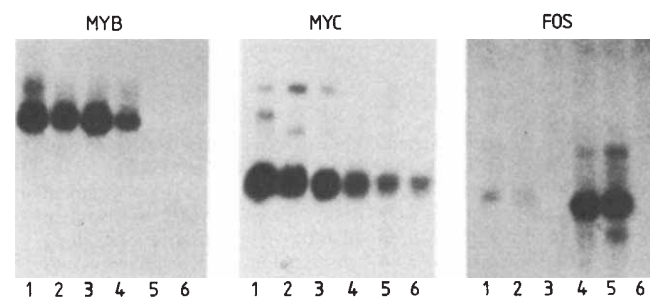


Fig. 4 Transcripts of *c-myc*, *c-myb* and *c-fos* genes in several cell types. Polyadenylated RNA was isolated from the following cell types: lanes 1, WEHI-3B; lanes 2, 416B¹⁵; lanes 3, B6.SutA.C1.15¹⁶; lanes 4, a low-density fraction of haematopoietic cells from 14-day murine fetal liver¹⁴; lanes 5, bone marrow-derived macrophages, which were the adherent cells obtained after culturing mouse bone marrow cells in the presence of macrophage-colony stimulating factor for 7 days³²; lanes 6, BALB/c 3T3 fibroblasts. All methods were as described in the legends to Figs 2 and 3.

Polymorphism and absence of Leu-enkephalin sequences in proenkephalin genes in *Xenopus laevis*

Gerard J. M. Martens & Edward Herbert

Department of Chemistry, University of Oregon, Eugene, Oregon 97403, USA

The structures of the genes coding for the opioid peptide precursors proopiometanocortin, proenkephalin (proenkephalin A) and prodynorphin (proenkephalin B), are known for some mammalian species^{1–7}. To gain insight into the evolutionary history of these precursors, we have examined the proenkephalin gene in the South African clawed toad, *Xenopus laevis*, which diverged from the principal line of vertebrate evolution some 350 Myr ago. The human proenkephalin gene consists of four exons, of which the main exon (exon IV) contains all known biologically active peptides—six Met-enkephalin sequences and one Leu-enkephalin sequence^{5,6}. We report here the primary structures of the putative main exons of two proenkephalin genes in *X. laevis*, each of which codes for seven Met-enkephalin sequences but no Leu-enkephalin, indicating that Met-enkephalin preceded Leu-enkephalin in the evolution of the proenkephalin gene. The organization of the main

exons of the toad genes is remarkably similar to that of the human gene and conserved regions provide evidence for functionally significant structures. We also detect a polymorphism in one of the toad proenkephalin genes, mapping 1.5 kilobases (kb) 5' of the main exon; it is caused by an insertion/deletion of a 1-kb repetitive sequence which has the characteristics of a transposable element.

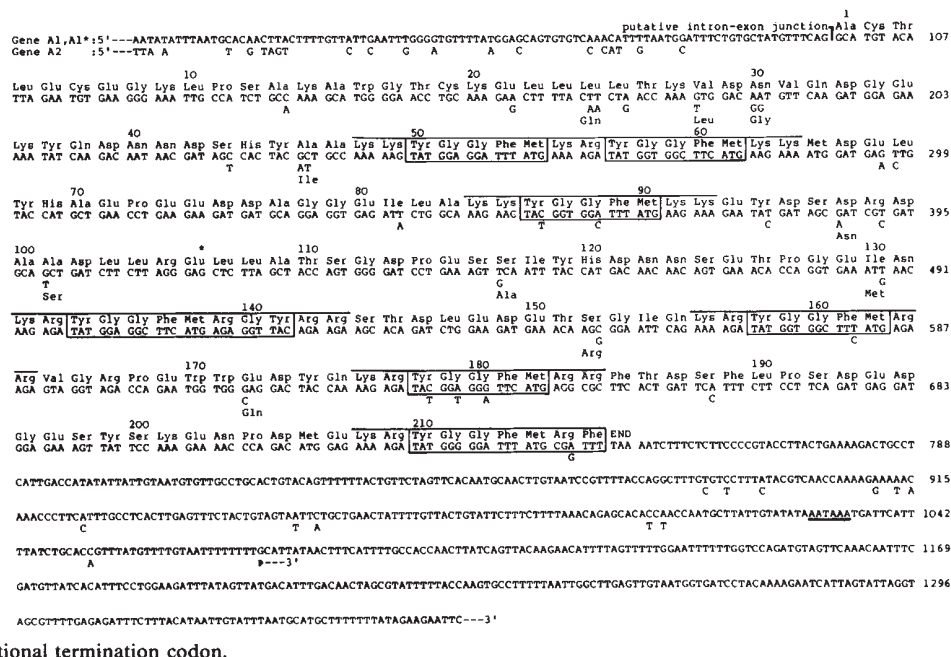
A *Xenopus* genomic library (kindly provided by Drs Y. Chien and I. B. Dawid) was screened by hybridization *in situ* with a human proenkephalin cDNA probe containing all of the protein-coding region. Two hybridization-positive phage clones could be isolated. A 930-base pair (bp) hybridizing fragment of the DNA insert carried by one of these clones was used as probe in a second screening of the library, which resulted in the isolation of six further hybridization-positive clones. Restriction maps of the eight clones showed them to comprise three different classes (λ XEA1, λ XEA1* and λ XEA2). Figure 1 shows nucleotide sequences, and deduced amino acid sequences, of the regions containing the hybridizing fragments of λ XEA1, λ XEA1* and λ XEA2, designated *Xenopus* proenkephalin genes A1, A1* and A2 respectively (see Fig. 1 legend for experimental details).

The putative 3' splice sites of the introns preceding the main exons of the *Xenopus* proenkephalin genes were tentatively assigned to A-G nucleotides 97 and 98 because these positions correspond to the splice site in the human counterpart^{5,6} and are preceded by regions resembling the splice junction acceptor consensus sequence in mammals⁸. In each *Xenopus* proenkephalin gene, the protein-coding region would then correspond to nucleotides 99–746, with nucleotides 747–749 (TAA) representing the first in-frame translational termination codons

(Fig. 1). Each of the main exons of the *Xenopus* genes thus codes for 216 amino acids which is slightly less than in human (221 amino acids). Toad proenkephalin is predicted to contain five copies of Met-enkephalin, one copy of Met-enkephalin-Arg-Gly-Tyr and one copy of Met-enkephalin-Arg-Phe. Human proenkephalin contains four copies of Met-enkephalin and one copy each of Met-enkephalin-Arg-Gly-Leu, Met-enkephalin-Arg-Phe and Leu-enkephalin^{5,6}. It is tempting, therefore, to speculate that the switch from a Met-enkephalin- to a Leu-enkephalin-coding sequence in the human gene occurred less than 350 Myr ago and that the Met-enkephalin sequence was the primordial duplication unit during evolution of the proenkephalin gene. In both species, the enkephalin sequences are flanked by pairs of basic amino acids, suggesting that similar processing mechanisms, involving trypsin-like enzymes, are necessary for production of the enkephalins. In each of the toad proenkephalin genes, only one polyadenylation signal consensus sequence⁹ is found (nucleotides 1,028–1,033); in the corresponding region of the human gene, four such signals are present⁶.

To examine the 5' ends of the *Xenopus* genes, DNA from clones λ XEA1, λ XEA1* and λ XEA2 was probed in non-stringent hybridization conditions with a human proenkephalin gene fragment containing exons I–III which code for the 5'-untranslated region and the N-terminal portion of human preproenkephalin. As no hybridizing fragment was detected in these experiments and in view of the lengths of the 5'-flanking regions of the main exons in the clones examined, it is likely that there is little homology between the 5' ends of the toad and human genes, although the possibility that these regions are absent from the clones cannot be excluded.

Fig. 1 Nucleotide sequences of the putative *Xenopus* proenkephalin genes A1, A1* and A2 (main exons and their flanking regions) and deduced amino acid sequences. The nucleotide and amino acid differences found in the A2 sequences are shown beneath the sequences of A1 and A1*; absence of a nucleotide or an amino acid in the A2 sequences indicates that the sequences of A1, A1* and A2 are the same at that position. The asterisk above nucleotide 416 indicates the only difference (G substituted by A in gene A1*); silent-site substitution between the presented nucleotide sequences of genes A1 and A1*. The arrowhead beneath nucleotide 1,077 indicates the 3' end of the sequenced portion of gene A2. The sequences of Met-enkephalin, Met-enkephalin with a carboxyl extension and their coding nucleotides are boxed; the pairs of basic amino acids flanking the enkephalins are overlined. The putative intron-exon junction is indicated by a vertical line; END, translational termination codon.



Methods: The *Xenopus* genomic library screened was a collection of recombinant phages that carried *Xenopus* blood-cell DNA fragments from several animals. The fragments were generated by partial *Mbo*I digestion of genomic DNA and cloned into the *Bam*HI sites of the vector λ J1. During the first screening of the library, the hybridization probe was a 918-bp *Hinc*II human proenkephalin cDNA fragment (which contains the full coding region) ³²P-labelled by nick-translation²⁴ to 5×10^8 c.p.m. per μ g DNA. Hybridization was carried out in $5 \times$ SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 1 mM EDTA, 0.06% each of bovine serum albumin, Ficoll and polyvinylpyrrolidone, 50 μ g ml⁻¹ of sheared and denatured herring sperm DNA and 10% dextran sulphate at 60 °C for 13 h. Two hybridization-positive phage clones (λ XE1–2) were isolated from about 7.5×10^5 plaques by repeated plaque purification. Partial restriction maps of these clones showed that they carried identical DNA inserts of 17 kb, containing two *Eco*RI fragments with approximate lengths of 2.0 and 0.8 kb which hybridized with the human cDNA probe. For further analysis, the two *Eco*RI fragments from λ XE1 DNA were subcloned into pBR322 and the subclones analysed by restriction mapping and DNA sequencing, resulting in the sequence of *Xenopus* proenkephalin gene A1. For the second library screening, the hybridization probe was a 930-bp *Hinc*II fragment of λ XE1 DNA (which contains the main exon of gene A1 and part of its 5'- and 3'-flanking regions) ³²P-labelled by random priming⁶ to 10^9 c.p.m. per μ g DNA and hybridization was at 65 °C for 13 h. Six hybridization-positive phage clones (λ XE3–8) were isolated from about 10^6 plaques. On restriction mapping, λ XE3–4 appeared to be identical to λ XE1–2, λ XE5–6 carried identical inserts of 18 kb and λ XE7–8 carried identical 19-kb inserts. The three types of clones were designated λ XEA1 (λ XE1–4), λ XEA1* (λ XE5–6) and λ XEA2 (λ XE7–8). The pertinent restriction fragments of λ XEA1* and λ XEA2 were subcloned into pBR322 and sequenced, resulting in the sequences of *Xenopus* proenkephalin genes A1* and A2 respectively. Sequencing was performed with the dideoxy chain-termination method²⁵ using M13 mp8 or mp9 RF-DNA as described previously²⁶.

A comparison of the predicted amino acid sequences of part of bovine, human and *Xenopus* proenkephalin (Fig. 2a) shows that the distribution of enkephalin sequences is remarkably similar among the three species. While several spacer regions share a high degree of amino acid homology, others have diverged considerably. Interestingly, the highly conserved regions between enkephalin units 2 and 3, 5 and 6, and 6 and 7 correspond to enkephalin-containing peptides isolated from bovine adrenal medulla¹⁰ (peptides F, E and B, respectively). The high degree of conservation of these peptides (especially of the highly potent opioid peptide E) might point to an important physiological role for them, both in mammals and lower vertebrates.

Met-enkephalin and Met-enkephalin-Arg-Phe, but little or no Met-enkephalin-Arg-Gly-Leu or Leu-enkephalin, have been

found in frog brain¹¹. These observations are consistent with our proposed amino acid sequence of toad proenkephalin which shows the absence of Met-enkephalin-Arg-Gly-Leu and Leu-enkephalin. Clearly, then, the proenkephalin gene(s) is expressed in amphibians but no function has so far been assigned to enkephalins in these organisms.

The regions of amino acid homology in the human and *Xenopus* proenkephalin sequences are reflected by regions of nucleotide sequence homology (Fig. 2). In addition, there are two homologous regions in the 3'-untranslated portions of the two genes. The conserved region preceding the polyadenylation signal might be involved in some post-transcriptional activity, as suggested for such a region in other genes^{12,13}; the significance of the other homologous region is unknown.

Fig. 2 Homology between *Xenopus* and mammalian proenkephalin sequences. *a*, Alignment of the amino acid sequences of part of bovine, human and *Xenopus* proenkephalin.

The one-letter amino acid notation is used. Residues identical among the three species are boxed. Gaps (-) have been introduced to achieve maximum homology. The locations of the enkephalin sequences are indicated by thick lines below the sequences. The bovine and human sequence data have been taken from refs 27, and 5 and 6 respectively. The *Xenopus* sequence is derived from *Xenopus* proenkephalin gene A1; similar conclusions as those given in the text can be drawn if the sequence derived from gene A2 were included. *b*, Schematic representation of the nucleotide sequence homology between the main exons plus their flanking regions of the *Xenopus* and human proenkephalin genes. The homologies were searched using the algorithms available on the SEQ computerized sequence analysis system from Stanford University, California. Deletions of nucleotides, introduced to achieve maximum homology, are shown as loops. The putative signal sites for polyadenylation (AATAAA), translational termination (TAA) and splicing (AG) are indicated. The locations of the enkephalin-coding sequences are indicated by lines below the schematic. The human sequence data were taken from refs 5 and 6. The *Xenopus* gene is gene A1; similar results are obtained when the human gene and *Xenopus* gene A2 are compared.

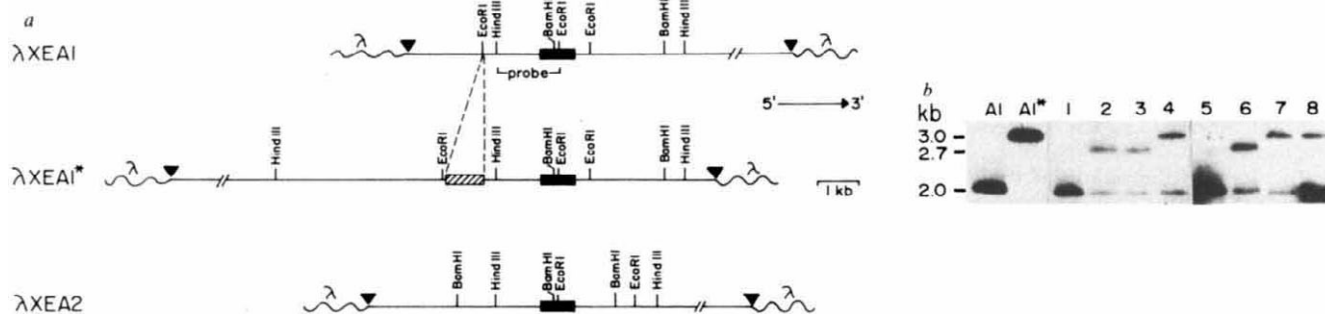
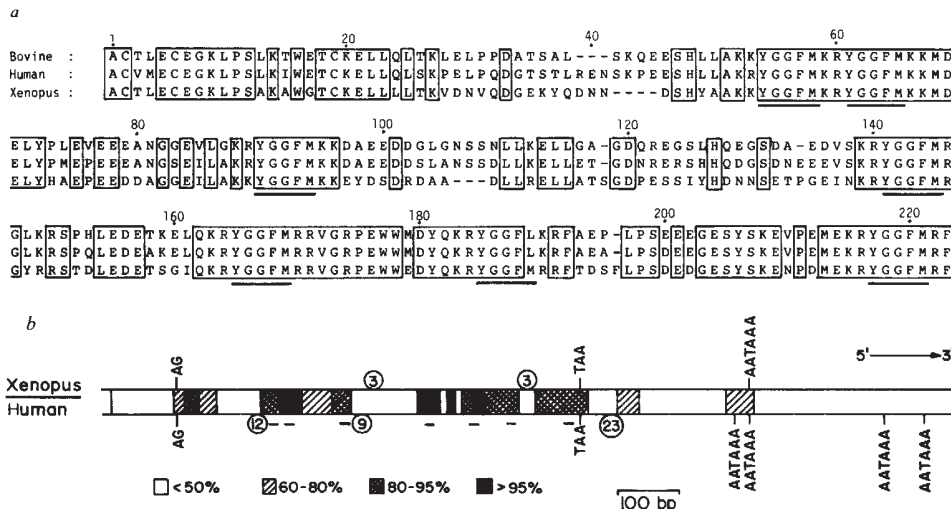
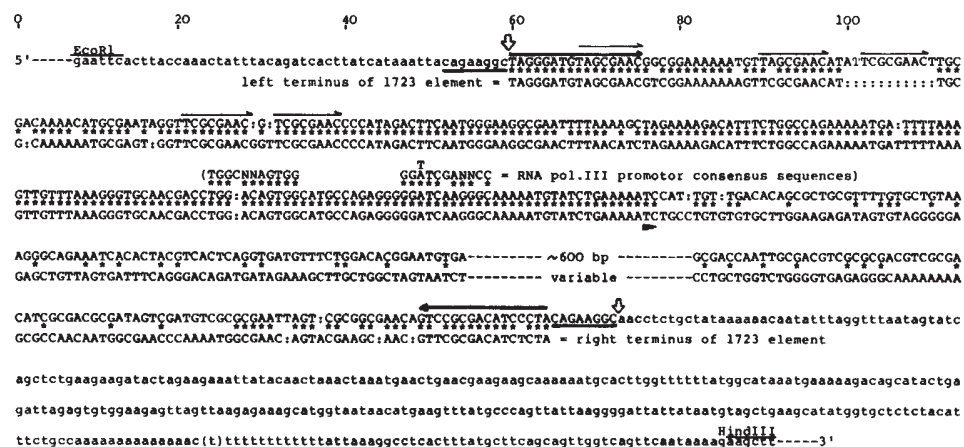


Fig. 3 Analysis of the polymorphism in *Xenopus* proenkephalin gene A1 by restriction mapping and Southern blotting. *a*, Partial restriction maps obtained from *Xenopus* proenkephalin genomic clones λXEA1 (17 kb), λXEA1* (18 kb) and λXEA2 (19 kb). For reference, the locations of the main exons of proenkephalin genes A1 (in λXEA1), A1* (in λXEA1*) and A2 (in λXEA2) are shown by closed boxes. The 1-kb DNA insertion ~1.5 kb 5' of the main exon of gene A1* is indicated by the hatched box in the map of λXEA1*. Arrowheads indicate the multiple cloning sites (BamHI, HindIII and EcoRI) of the vector λJ1; wavy lines indicate the arms of λJ1. The region covered by the probe used in the Southern blot analysis shown in *b* is indicated below the map of λXEA1. *b*, Southern blot analysis of EcoRI digests of *Xenopus* liver DNA isolated from eight *X. laevis* individuals (lanes 1-8) and of EcoRI digests of DNA from genomic clones λXEA1 and λXEA1* (lanes A1 and A1* respectively). After digestion of 10 µg high-molecular weight genomic DNA with EcoRI, the restriction fragments were separated on a 0.9% agarose gel. The fragments were transferred to a nitrocellulose filter²⁸, the filter was probed with the 1.6-kb EcoRI-HindIII fragment of λXEA1 DNA (see under *a*) which was subcloned into M13 and ³²P-labelled by primer extension to 5 × 10⁸ c.p.m. per µg DNA, and the blot was autoradiographed with an intensifying screen at -70 °C. The hybridizing 2.7-kb EcoRI fragment found only in animals 2, 3 and 6 did not correspond to any of the EcoRI fragments of the isolated genomic clones, so its identity remains unclear. The remainder of the blot (not presented for clarity) only showed an approximately 14-kb EcoRI fragment found in all animals, which suggests that the hybridizing 4.9-kb EcoRI fragment of clone λXEA2 (see also the restriction map of λXEA2 in *a*) is generated by cleavage at the EcoRI site within the main exon of gene A2 and at the EcoRI site in the multiple cloning site of λJ1.

Fig. 4 Partial nucleotide sequence of the insert DNA and its flanking regions. The 0.4-kb *EcoRI*–*HindIII* fragment and the ends of the 1.4-kb *EcoRI*–*HindIII* fragment located 5' of the main exons of *Xenopus* proenkephalin genes A1 and A1* in λ XEA1 and λ XEA1* respectively (see also Fig. 3a) were subcloned into mp8 and mp9 M13 RF-DNA and sequenced with the dideoxy chain-termination method²⁵ as described previously²⁶. Lower-case letters give the sequence common to both the 0.4- and the 1.4-kb fragments (except for the presence of an additional T nucleotide 63 bp from the *HindIII* site in the 1.4-kb fragment); the sequenced portion of the insert DNA is in upper-case letters.

Open arrows indicate the positions of insertion. The direct repeats flanking the insert are underlined. The inverted nucleotide sequences at the ends of the insert are overlined by heavy arrows; the part of the inverted nucleotide sequence which is repeated several times in the 5' end of the insert (T₅GCGAAC) is overlined by light arrows. The nucleotide sequences of the left and right termini of the 1723 element, taken from ref. 22, are shown beneath the sequence of the insert. Nucleotides identical between the inserted sequence and the 1723 element are indicated by asterisks. Colons indicate gaps introduced to maximize homology. The arrowhead indicates the start (5' end) of an approximately 180-bp inverted nucleotide sequence in the 1723 element²². The putative consensus sequences for an RNA polymerase III promoter^{20,21} (N is any nucleotide) are given above the sequence of the insert and they have been aligned to show homology.



The nucleotide sequences of the main exons of *Xenopus* proenkephalin genes A1 and A2 differ by ~4.6% and this sequence divergence is similar to that reported for the exons of other known *X. laevis* gene pairs coding for globins¹⁴, albumins¹⁵ and vitellogenins¹⁶ (5–8%). It is thought that the gene pairs represent duplicated genes resulting from a genome duplication in the genus *Xenopus* some 30 Myr ago, which was originally postulated on the basis of chromosomal studies^{17,18}.

Not only the nucleotide sequences of the main exons of *Xenopus* proenkephalin genes A1 and A1* but also their flanking regions are virtually identical (Figs 1, 3a). The only difference detectable in the restriction maps of these two sequences is the location of an *EcoRI* site 5' of the main exons (see Fig. 3a), which suggests that gene A1* is an allelic variation of gene A1. Figure 3b shows the distribution of the polymorphism in *Xenopus* proenkephalin gene A1. Three of the animals examined were heterozygous at this locus and two animals were homozygous for gene A1. We could not determine the relationship between the hybridizing 2.7-kb *EcoRI* fragment found in animals 2, 3 and 6 (Fig. 3b) and genes A1 and A1* because despite extensive screening of the *Xenopus* genomic library, we could not isolate a clone containing the 2.7-kb fragment. We conclude that the library was constructed from animals lacking this fragment.

To establish the nature of the allelic variation of gene A1 (gene A1*), the 0.4-kb and the ends of the 1.4-kb *EcoRI*–*HindIII* fragments of λ XEA1 and λ XEA1*, respectively, were sequenced. The polymorphism seems to be caused by a 1-kb DNA insertion, approximately 1.5 kb 5' of the main exon of gene A1 (Figs 3a, 4). Since the 5' ends of the *Xenopus* proenkephalin genes are not yet known, the positioning of the insert within the gene remains unclear. The insert is characterized by nearly perfect 16-bp inverted nucleotide sequences at its ends (part of which, T₅GCGAAC, is repeated several times) and is flanked by 8-bp direct repeats, of which one may have been generated by target site duplication during integration of the element (Fig. 4). Thus, the insert has the characteristics of a transposable element¹⁹. No long open translational reading frame was found in the sequenced portion of the insert. The two boxes that share homology with the putative RNA polymerase III promoter consensus sequences^{20,21} are unusually close to one another (Fig. 4). Using hybridization it was shown that the insert DNA is highly repetitive and is present in about 5,000 copies per haploid genome (data not shown). The insert may belong to the family of 1723 elements, a recently described group of repetitive

sequences interspersed in the genome of *X. laevis*²², since a region in the 5' end of the insert (comprising more than 200 nucleotides) and the inverted repeats at its ends are highly homologous to the left terminus and the inverted repeats, respectively, of the 1723 elements. As transcripts of the repetitive 1723 elements are found among *Xenopus* tadpole polyadenylated RNAs²² and repetitive sequence transcripts have been implicated in the regulation of gene expression during ontogeny²³, it is interesting that so far we have only found homozygotes for proenkephalin gene A1* in tadpoles and never in adult animals. Thus, we are interested in the possibility that the insertion/deletion of the transposable-like repetitive element in the close vicinity of proenkephalin gene A1 is involved in the expression of this gene during *Xenopus* development.

We thank Michael Comb for providing the human proenkephalin cDNA and for helpful discussions. This work was supported by NIH grants AM16879, AM30155 and DA02736 (E.H. Principal Investigator). G.J.M.M. is supported by a postdoctoral fellowship from the Netherlands Organization for the Advancement of Pure Research (ZWO).

Received 1 February; accepted 24 May 1984.

- Nakanishi, S. *et al. Eur. J. Biochem.* **115**, 429–438 (1981).
- Cochet, M., Chang, A. C. Y. & Cohen, S. N. *Nature* **297**, 335–339 (1982).
- Whitfield, P. L., Seeburg, P. H. & Shine, J. *J. DNA* **1**, 133–143 (1982).
- Notake, M. *et al. FEBS Lett.* **156**, 67–71 (1983).
- Noda, M. *et al. Nature* **297**, 431–434 (1982).
- Comb, M., Rosen, H., Seeburg, P. H., Adelman, J. & Herbert, E. *DNA* **2**, 213–229 (1983).
- Horikawa, S. *et al. Nature* **306**, 611–614 (1983).
- Mount, S. M. *Nucleic Acids Res.* **10**, 459–472 (1982).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211–214 (1976).
- Kilpatrick, D. L. *et al. Proc. natn. Acad. Sci. U.S.A.* **78**, 3265–3268 (1981).
- Kilpatrick, D. L., Howells, R. D., Lahm, H. W. & Udenfriend, S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5772–5775 (1983).
- Van Ooyen, A., van den Berg, J., Mantel, N. & Weismann, C. *Science* **206**, 337–344 (1979).
- Ordahl, C. P. & Cooper, T. A. *Nature* **303**, 348–349 (1983).
- Widmer, H. J., Andres, A. C., Niessing, J., Hoshbach, H. A. & Weber, R. *Dev. Biol.* **88**, 325–332 (1981).
- May, F. E. B., Westley, B. R., Wyler, T. & Weber, R. *J. molec. Biol.* **168**, 229–249 (1983).
- Germond, J. E. *et al. Nucleic Acids Res.* **11**, 2979–2997 (1983).
- Bisbee, C. A., Baker, M. A., Wilson, A. C., Hadji-Azimi, I. & Fischberg, M. *Science* **195**, 785–787 (1977).
- Thiébaud, C. H. & Fischberg, M. *Chromosoma* **59**, 253–257 (1977).
- Kleckner, N. A. *Rev. Genet.* **15**, 341–404 (1981).
- Galli, G., Hofstetter, H. & Birnstiel, M. L. *Nature* **294**, 626–631 (1981).
- Traboni, C., Ciliberto, G. & Cortese, R. *EMBO J.* **1**, 415–420 (1982).
- Kay, B. K. & Dawid, I. B. *J. molec. Biol.* **170**, 583–596 (1983).
- Davidson, E. H. & Britten, R. J. *Science* **204**, 1052–1059 (1979).
- Weinstock, R., Sweet, R., Weiss, M., Cedar, H. & Axel, R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1299–1303 (1978).
- Sanger, F., Nickle, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Comb, M., Seeburg, P. H., Adelman, J., Eiden, L. & Herbert, E. *Nature* **295**, 663–666 (1982).
- Noda, M. *et al. Nature* **295**, 202–206 (1982).
- Southern, E. M. *J. molec. Biol.* **98**, 503–517 (1975).

A view from the left

P.B. Medawar

Not in Our Genes: Biology, Ideology, and Human Nature.

By R.C. Lewontin, Steven Rose and Leon J. Kamin.

Pantheon, New York: 1984. Pp.322. \$21.95. To be published in Britain on 27 September by Penguin, pbk £3.95.

THIS book by three extremely intelligent and articulate writers — “respectively an evolutionary biologist, a neurobiologist and a psychologist” — is a polished and in the main convincing rebuttal of a variety of determinist ideologies that have come to acquire the status of a public nuisance in biology and sociology. The two most offensive of these are widely known by the names of “historicism” (Popper’s word, I think) and “geneticism” (I think mine).

Geneticism is the doctrine of the primacy of the genetic make-up in determining every aspect of the human mind, constitution and social behaviour: we must look to genetics for an explanation of tribalism and aggression, the rise and fall of nations, the stratification of societies into classes, the nature and degree of human intellectual prowess and for the ability of some English-speaking people to pronounce the consonant theta or the inability of others (continentals generally, I fear) to do so. Geneticism systematically depreciates — indeed, resents — the existence of the kind of heredity (exosomatic) that is mediated not through chromosomes but through cultural means. Geneticism is most tiresome and mischievous in the context of what I compendiously call the “IQ nonsense”, resting as it does on a two-fold illusion: that a complex quantity can be rendered by a single figure (as a man’s height can be, or his body temperature) and that it is possible to allocate percentage figures to the relative contributions of nature and nurture to the development of any trait whatsoever. This topic was the subject of the silliest remark I have ever heard during what has admittedly been a sheltered and unworldly academic life: I refer to Professor H.J. Eysenck’s judgement — on p.111 of *The Inequality of Man* (Temple Smith, 1973) — that “the whole course of development of a child’s intellectual capabilities is largely laid down genetically”.

Because this present work is so skilfully and successfully presented as the work of a team, the IQ nonsense in its various aspects is not dealt with as a separate chapter credited to Leon Kamin; but his is naturally the name we think of because it was he who first drew public attention to the political motivations that underlie the whole nonsense and who played a principal part in uncovering the frauds of Sir Cyril Burt (which were finally laid bare by an investigative journalist with genetic training, Dr Oliver J. Gillie). The Burt deceptions

reflect not merely upon their vain and mischievous old author but on the psychological profession as a whole, for how on earth did Burt get away with it for so long and what about all that fine talk of the avoidance of fraud through the self-monitoring of scientific research? The reason is so obvious when pointed out that it is often not pointed out: the IQ boys had no incentive or inclination to scrutinize Burt’s results carefully, *because he told them exactly what they wanted to hear* — and I suppose Burt’s having been knighted was thought vicariously to confer upon them the degree of respectability they were, as they still are, very much in need of. The whole of Chapter 5 of *Not in Our Genes*, of which Sir Cyril Burt is protagonist, struck me as sound stuff and moreover is good, lively reading — indeed this book is nowhere a book to be nodded over except perhaps in assent.

The authors see the various manifestations of the determinism they disapprove of as so many manifestations of Tory ideology; they are surely right, for is not it a canon of high Tory philosophy that (in my own words) “a man’s *breeding* determines absolutely his capabilities, his destiny and his deserts”. Certainly Tory philosophy is utterly incompatible with the authors’ commitment to the possibility of creating “a more sociologically just — a socialist — society” (p.ix).

The political incentive to cultivate geneticism would be hard to rebut, associated as it is with the more virulent kind of racism — that which imputes an inborn superiority to some people and inferiority to others. The case for a Tory political motivation of historicism is however less convincing.

Historicism is the doctrine that there exists or can be propounded an historical social science that makes possible ostensibly scientific predictions about the future course of history and the future condition of man. The principal context of historicism is of course sociology and the study of history, but there are strongly historicist as well as geneticist elements in sociobiology, for which Lewontin, Rose and Kamin have a special repugnance; and they trace the origins of its popular appeal in what Stephen Gould once described as “pop ethology” as it is exemplified by the writings of Robert Ardrey and Desmond Morris, seeing E.O. Wilson as a sort of modern Thomas Hobbes. “Sociobiology is Pangloss made scientific through the agency of Charles Darwin” — Pangloss,

one of the principal characters of Voltaire’s tiresome *Candide*, is he who contended that all was for the best in what must of necessity be the best of all possible worlds. It is not among professionals, but educated men who think that they are having terribly deep thoughts when they read Robert Ardrey and the like that sociobiology has its greatest fan following: this is all the more surprising because I cannot call to mind and have never been informed of any genuinely novel or illuminating insight into the human condition that has come to us through sociobiology.

By contrast with geneticism, historicism is not the outcome of Tory political thinking; indeed, the first and certainly the most influential historicist thinker was Karl Marx whose economic interpretation of history led to sociohistorical predictions having to do with the inevitable deterioration of the workers’ lot, for example — predictions which history has not borne out.

I liked the content of this book, most of which seemed to me to have a “right-thinking” quality. I liked also the skill with which the multiple-authorship has been concealed, and the punchy, direct style wholly free from quasi-philosophical reservations and prevarications.

The only important point on which I disagree with the authors is their tendency to treat reductionism as an “ism” as wrong-headed as the two I have been discussing above. Opposition to reductionism is now a popular posture among sociologists and people with a rather literary outlook on science. The attack, I believe, was launched by a number of failed scientists such as the late Ludwig von Bertalanffy to avenge themselves against a science in which they had failed to become proficient by creating a travesty of the concept to bring it as far as possible into discredit. It was Bertalanffy, for example, who invented the analytical-summative-mechanist who believes a whole is no more than the sum of its parts and that its properties are fully explicable in terms of the properties of its parts, having no regard to the functional relationships existing between them. Whenever I read denunciations of these ghoulish figures I wait impatiently for the identity of at least one of them to be revealed so that with pious repugnance I can study their writings for myself. But its practitioners are never named, simply because they do not exist.

Reductionism has philosophic weaknesses (what methodology has not?). Claims for its efficacy such as John Stuart Mill made are very vulnerable to *reductio ad absurdum* and are not taken seriously. Thus “being interpretable in terms of” is what logicians call a “transitive” relationship, for if A is interpretable in terms of B and B in terms of C then A must be interpretable in terms of C. Thus a concept at the socio-political level such as that which is embodied in the Married Women’s

Property Act (1882) should be interpretable in terms of physics and chemistry. If it is so, it is strange that it does not appear in the syllabus of physics in any educational establishment known to them, nor, for that matter, does proportional representation, nor (my favourite example for the purpose) the foreign exchange deficit; but in spite of these shortcomings reductive analysis is the most successful research stratagem ever devised in science. What is more — something that should appeal to left-wingers such as those who wrote this book — it is that way of understanding the world which makes it easiest to see how, if need be, the world can be changed, something which, like myself, the authors of this book do indeed believe it to be in need of. □

Sir Peter Medawar is at the Medical Research Council's Clinical Research Centre, Harrow, Middlesex. His most recent book, written with Jean Medawar, is Aristotle to Zoos: A Philosophical Dictionary of Biology (Weidenfeld & Nicolson/Harvard University Press, 1984).

Neuropeptidology

M. Ian Phillips

Brain Peptides.

Edited by Dorothy T. Krieger, Michael J. Brownstein and Joseph B. Martin.
Wiley: 1984. Pp. 1,032. \$112.15, £83.10.

WHEN a field of research opens up, there is often a rush to stake-out territory reminiscent of the opening of new frontiers in the Wild West. Such has been the case with the study of neuropeptides. What has made the subject exciting is that so many peptides appear to exist in the brain that were not known or even suspected a few years ago. All kinds of claims have resulted, both substantial and exaggerated.

The editors of *Brain Peptides* are at home on this range, and have decided that it was time to sit back on the porch and take a look at all the activity. To do this, they have invited some of the leading pioneers, and a number of recent settlers, to write chapters on specific topics.

There are excellent contributions on peptide hormone genes, biosynthesis of neuropeptides and aspects of peptides in cell biology, evolution and embryology written by Miller, Loh and Gainer, McKelvy, Acher, Strumwasser and Price. The role of peptides in homeostatic systems is covered by several authors, including Watson and Akil, Brown, Koob and Bloom, Pfaff, Fink and Karten. However the choice of Reid to write an article on

angiotensin in a book about brain peptides is ironic, since he argues that angiotensin does not exist in the brain.

A particularly useful section of the book is that dealing with methodology. The technique of immunocytochemistry is a clear favourite among the contributors for mapping and surveying the new areas — Palkovits's chapter is excellent on this and other neuroanatomical techniques. Many chapters contain lists of sites in the brain which have the ubiquitous brown stains of immunoreactive material. Less frequently, these are backed up by radioimmunoassay and further identification with high-pressure liquid chromatography. The retina has also turned out to be a rich seam of several peptide-like substances. Brecha and Karten describe how they are localized in distinct retinal cell types and that they have been found in every species studied so far. Another site that is a veritable gold mine is the paraventricular nucleus where about 30 different peptides have been identified.

The section on specific peptides is written by several pioneers, Zimmerman, Mutt, Martin, Leeman, Dockray and Vale among them. One might think that each chapter would have the same type of information for each peptide. Oddly enough, this is not so; with some peptides it has been considered crucial to demonstrate synthesis, distribution, receptor binding, physiology and function, but with others it has not. Inevitably, progress has been faster in one direction with one peptide than with another. In part this reflects the difficulties encountered with certain peptides, in part the interest and enthusiasm of an individual investigator.

The material contained in *Brain Peptides* was previously scattered through various symposium proceedings or review articles, but while the book offers a handy source it is not a work of reference. Some authors have included only skimpy reference lists, while those of others — such as Liotta and Krieger — are all-inclusive. Another complaint is that references are given in abbreviated form without titles, which probably saves space in the book but costs time in the library for the reader.

Reading *Brain Peptides* makes one realize not so much what is new but what answers are still required. For example, more detailed characterization of the immunoreactive substances is necessary, as are more electronmicroscopic analyses. A theoretical basis for peptide action is elusive and much needed; the statement found here that peptides have a "modulatory action" explains nothing. Nevertheless this book is a timely stopping point for reflection and for survey of a new field to explore — not strictly of brain peptides, but the study of neuropeptides which might be called "neuropeptidology". □

M. Ian Phillips is Professor in and Chairman of the Department of Physiology, University of Florida.

Multiple uses for monoclonals

Brian Anderton

Monoclonal Antibodies.

By Karol Sikora and Howard M. Smedley.

Blackwell Scientific: 1984. Pp. 132.
Pbk £8.50, \$16.75.

MONOCLONAL antibodies have proved to be powerful tools in diverse areas of biology and medicine. There are probably now many more non-immunologists than immunologists using them, a fact which makes short, easily readable publications such as this very valuable.

In the preface, the authors state that their aim was to produce a book on monoclonal antibodies aimed at students of medicine as well as physicians. In this they have succeeded exceedingly well, but it is unfortunate that the title does not indicate the bias. There is only one chapter devoted to the application of monoclonal antibodies in cell biology, but its treatment of the topic is so slight that the value of the book to biologists is far less than that to medical students. This is a pity because the first three chapters on the basics of antibody properties, production and assay are written at a level such that a prior grounding in immunology is not essential. Also, while the book includes a number of descriptions of practical procedures, it does not set out to be a laboratory manual of detailed protocols.

The meat of the text is a comprehensive discussion of monoclonal antibodies in medicine, where they are already proven and are increasingly used routinely for a variety of diagnostic tests in all of the pathology services. The authors also point out that the future will see further antibody-based tests, replacing slow and less precise methods (for example in identification of certain microorganisms), and the introduction of new tests (say, for head injury, by the immunoassay of brain antigens in serum). Less certain is the potential for the use of monoclonal antibodies in therapy, particularly in cancer patients. This topic is of special interest to the authors and although they present an optimistic view, the subject is treated in an objective way. Their even-handed approach also applies to the problem of making human monoclonal antibodies, the solving of which remains a highly desirable goal.

Sikora and Smedley's book is a reasonably priced and, of course, a timely publication. It should be read by all those training or working in medical science who are not already conversant with monoclonal antibodies, be they sceptics or not. □

Brian Anderton is Senior Lecturer in Immunology at St George's Hospital Medical School, University of London.

New in paperback

The Clocks That Time Us: Physiology of the Circadian Timing System, by M. C. Moore-Ede, F. M. Sulzman and C. A. Fuller. Publisher is Harvard University Press, price is \$10.95, £9.30. For review see *Nature* 299, 323; 1982.

Between the bedside and the laboratory

D.J. Weatherall

The Encyclopaedia of Medical Ignorance: Exploring the Frontiers of Medical Knowledge.

Edited by Ronald Duncan and Miranda Weston-Smith.

Pergamon: 1984. Pp.253. £9.95, \$17.95.

RECENTLY, I took a group of new clinical students to see a patient suffering from the severe bone pain of sickle cell anaemia. Here, surely, was encapsulated everything that two or three years rigorous training in the basic sciences had prepared them for.

The teaching session started well enough. A single base change in one of the patient's globin genes had led to the production of an abnormal haemoglobin which, in turn, had caused his red cells to change their shape. This had led to blockage of the microcirculation, death of tissue, and hence to the severe pain from which he was complaining. But then the awkward questions started. Why did his red cells sickle and produce these symptoms today; he was playing football yesterday? Why has his brother with the same disease never had any pain? How are you going to treat him and prevent a recurrence of the pain? And so on.

Much of modern medicine is like this; for a few diseases we have an inkling about what may be happening at the cellular or molecular level, or at least we can start to understand the ailment in terms of whole-organ dysfunction. But it is still difficult to translate much of this knowledge into a rational explanation for what we see at the bedside, while for many of the common degenerative or neoplastic disorders that fill the wards we have not yet even reached this stage. I suspect that it is becoming increasingly difficult for our brighter students to compromise between the empiricism of good bedside practice and the relatively thin scientific evidence on which so much of it is based; as Lewis Thomas has pointed out, practitioners of clinical science have progressed from a state of total ignorance to a position in which they are aware of their ignorance.

It follows that all textbooks of medicine are to some extent compendia of ignorance. So why another, albeit under an explicit title — we already know we are ignorant. The editors' stated aim is to highlight some major lacunae in our knowledge, and in doing so to encourage fresh thought about how to tackle important areas of clinical uncertainty. Have they succeeded? A glance through the index suggested that I would not find much that I was unaware that I was ignorant about. About half of the book is devoted to psychiatry and neurology, and most of the remaining chapters cover equally pre-

dictable subjects such as blood pressure, cancer therapy, the immunology of pregnancy, transplantation biology and the fascinating mystery of how parasites manage to survive in their hosts.

There are some highlights and surprises, however. John Edwards writes a characteristically lively account of human mutation, and Leslie Iversen and Anthony Allison give excellent reviews of the present state of our understanding of neurochemistry and immunity to viruses. Philip Gell, in a thoughtful if rather despairing essay on the complexities of the human phenotype, seems to have little faith in the ability of the tools of modern biology to tell us anything useful about behaviour or free will — this



The Victorian way of death — Lady with the Tree Cross in Highgate Cemetery, North London. The picture is taken from Highgate Cemetery: Victorian Valhalla, a haunting celebration of the cemetery in words (by Felix Barker) and pictures (by John Gay). The book is published by John Murray, price £7.50.

chapter should be essential reading for television producers before their next venture into the gene supermarket. The only author who would have us believe that medical ignorance is bliss is Richard Peto, who suggests that our current obsession with the mechanistic approach to cancer research is diverting our efforts (and finances) from more obvious lines of epidemiological study. Why, he asks, do we want to understand the cell biology of malignant transformation when most human cancers are due to avoidable carcinogens. Should there be a second edition, I suspect that someone will tell him.

Overall, the essays in this attractively produced collection are of a high calibre and anybody interested in the state of clinical science will find something to engage them. Where I found the book rather lacking was in some of the broader

and particularly worrying problems of modern medicine. The World Health Organization has suggested that we aim for good health for all by the end of this century. But what is good health? By changing diet, stopping smoking, controlling blood pressure and pounding the pavements at the crack of dawn, the people of Western societies may partly control the major killers of middle life. But how will we cope with our increasingly aged population? We know nothing about the biology of ageing or about what the quality of life will be for our clean-living octogenarians. And what price will they pay for their longevity? There is the prospect that we will end up with institutions full of ninety-year-olds reflecting gloomily on years of missed opportunities caused by impotence and other side effects of the drugs that controlled their mild hypertension. And there are the possible effects of our current obsession with exercise in middle life on the later incidence of degenerative joint disease. Will we all end up with a healthy myocardium but totally immobile? At the other end of the age spectrum is the question of how far we are willing to go in the application of the techniques of molecular biology to the prevention and correction of genetic disease and congenital malformation. As genetic manipulation becomes feasible will positive eugenics once more become a topic of hot debate?

Perhaps most important, who is going to pay for all this, and pay for what — do our priorities lie in trying to understand the mechanisms of degenerative arterial disease and cancer, or in prevention, or in increasingly high technology patch-up medicine for the middle aged, or in providing facilities for the massive geriatric population that is being created? And who will determine these priorities? The medical profession is happy to go on developing expensive aids to survival, and the public seem to expect them. Politicians and those who administrate our health services should take note of the recent announcement of the successful installation of an artificial heart; the doctors of those who enjoy their food and drink, and who prefer to be in bed rather than plodding through the streets at sunrise, are already reaching for their order forms.

Equally pressing are the medical problems of the developing countries, virtually ignored in the book. As Maurice King has pointed out, 600 million people will be trapped in absolute poverty by the year 2000; that is, they will be living under conditions so characterized by malnutrition, illiteracy, disease, high infant mortality and low life expectancy as to be beneath any reasonable definition of human decency. The editors of this book should take heart; there is plenty of scope for future editions. □

D.J. Weatherall is Nuffield Professor of Clinical Medicine at the University of Oxford.

Reactions in theory

P.E. Hodgson

Direct Nuclear Reactions.

By G.R. Satchler.

Oxford University Press: 1983. Pp.833. £55, \$110.

Direct Nuclear Reactions.

By Norman K. Glendenning.

Academic: 1983. Pp.378. \$62, £45.50.

FOR the past 20 years the names of Ray Satchler and Norman Glendenning have been familiar to all students of nuclear reactions. These two men pioneered the development of practically every aspect of the theoretical formalism widely used to analyse experimental data and to extract information on nuclear structure. Both, inevitably, have been invited to write books. Ray Satchler first agreed to write his in 1960; Norman Glendenning began thinking about it in 1969 but did not start writing until 1978. And now the two books are published together.

Direct nuclear reactions are those processes that take place in the initial stages of a nuclear collision, in the time it takes the projectile to traverse the field of the target nucleus. Rather few degrees of freedom are excited by such processes, so that studies of the energies and angular distributions of the outgoing particles yield important information on nuclear structure. Subsequently, other processes take place, often with further particle emission.

Broadly speaking, there are three types of direct reactions — elastic scattering that leaves the target nucleus unchanged; inelastic scattering that raises it to an excited state; and transfer reactions that produce a different nucleus. Inelastic scattering tends to excite collective states in which many nucleons move coherently, while transfer reactions preferentially excite states that can be described as the excitation of one nucleon or group of nucleons. Thus different types of reactions can be used to explore very different types of nuclear excitation.

The physical ideas underlying our current understanding of these reaction processes are mostly quite simple, but their quantum mechanical formulation is frequently a task of formidable complexity and the resulting equations can only be solved by powerful computers. Both authors are theoreticians, with complete mastery of the formal techniques of reaction analysis and the angular momentum algebra needed in realistic calculations. So it is no surprise that the main part of each book is devoted to accounts of the formalisms appropriate to different types of direct reactions, and the approximations necessary to obtain reasonably accurate results in finite computing time.

There is inevitably much common ground between the two books, but with differences of emphasis and depth of treat-

ment. Both cover the basic reaction formalism, starting with the general expression for the transition amplitude, partial wave analysis and the optical model. Next in order of complexity comes the distorted wave theory of one-nucleon transfer reactions. Both books cover this in considerable detail, going on to describe the coupled-channel formalism that makes it possible to include higher order processes that are important when the coupling between channels is strong. This is essential to the understanding of the excitation of collective states by many different projectiles. The same technique may be applied to two-nucleon and multi-nucleon transfer reactions, with considerable increase in formal complexity. It is also possible to develop microscopic theories of the reactions, in which the participating nucleons are considered separately. Finally, the collisions of nuclei with each other, known as heavy-ion interactions, reveal a wealth of new phenomena that may be understood by extending the concepts and formalism developed for the interactions of the lighter particles.

On the whole, both books start from a level appropriate to the graduate student. Glendenning keeps to essentials, while Satchler describes many more detailed applications and special techniques. Both authors give many references to the original literature, and illustrate their discussions by comparisons with experimental data.

The scope of both books is almost entirely confined to direct reactions, with little mention of the subsequent pre-equilibrium and compound nucleus processes, or of their relative importance at different energies and for different projectiles and

target nuclei. The experimentalist is however confronted with a cross-section, and the first question he has to answer is which of these processes has occurred. Only when he knows the answer, at least provisionally, can he begin to analyse the data using the appropriate theory. The usefulness of the books would have been somewhat increased by more consideration of this problem.

A more subtle question encountered in the analysis of data is the degree of flexibility of the theories used. It is often relatively easy, by judicious parameter adjustment, to fit a theory to a known cross-section, but it is much harder to predict with confidence the value of a cross-section *before* it has been measured. It is notoriously difficult to judge the flexibility of a theory just by reading other people's papers — one has to repeat the analysis for oneself. Both authors are familiar with this problem, and it would have been valuable to learn more about their own experiences. One final — and minor — comment; selectivity is inevitable in works such as these, but in both books more references to peripheral areas of research would have been welcome, especially to proceedings of conferences and summer schools.

Satchler and Glendenning have undoubtedly earned the gratitude of nuclear physicists, and their books will be essential for reference and study. These volumes should be in every science library, and research workers would do well to save up until they can afford their own copies. □

P.E. Hodgson is Head of the Nuclear Physics Theoretical Group at the Nuclear Physics Laboratory, University of Oxford.

Diffusing knowledge

Charles DeLisi

Random Walks in Biology.

By Howard C. Berg.

Princeton University Press: 1984. Pp.142. \$16.50, £15.30.

A GOOD swimmer can typically coast about 18 ft, approximately three body lengths. A bacterium, on the other hand, once it stops propelling itself, will come to rest in 0.4Å — almost two orders of magnitude less than the diameter of a hydrogen atom. The inertial equivalent of a bacterium moving through water, is *Homo sapiens* swimming in butter! Evidently, an intuition about physical processes that occur on scales characteristic of our daily environment does not always extrapolate readily to the world of the very small. The purpose of Howard Berg's book is to sharpen our intuition about the physics and statistics that govern motion in the microscopic environment of the cell.

The random molecular motions of the microscopic world that manifest them-

selves by macroscopic regularity dominate both the motion of cells and the techniques used to study them. It is well known that the regular movement of a sedimentation band finds its microscopic explanation in the random movement of individual molecules. The effect of Brownian motion on the ability of a cell to capture efficiently ligands in its environment is, however, less well known. Who would guess that with only a fraction of a per cent of a cell surface covered by receptors, the rate at which ligands are captured can be nearly half that of a uniformly absorbing sphere!

Random Walks in Biology, which supplements and elaborates on the now classic paper by Berg and Edward Purcell ("Physics of Chemoreception" — *Biophys. J.* **20**, 193–219; 1977), is a scholarly and pedagogically masterly introduction to diffusion, its physics and its statistics. The book will probably require several readings in order fully to assimilate its concepts, but the time will be well and enjoyably spent. □

Charles DeLisi is Chief of the Laboratory of Mathematical Biology, National Institutes of Health, Bethesda, Maryland.

Awards

Upon approval by Her Majesty the Queen the Royal Society UK, has awarded three Royal Medals for 1984: to **Professor A.R. Battersby**, Professor of Organic Chemistry at the University of Cambridge, for contributions to the elucidation of the pathway for the biosynthesis of complex natural products; to **Dr Mary F. Lyon**, Member of the Scientific Staff of the Medical Research Council for contributions to the discovery of X-chromosome inactivation as a mechanism of gene dosage compensation; and to **Professor A.L. Cullen**, Emeritus Professor of Electrical Engineering at the University of London, for contributions to microwave engineering, both theoretical and experimental, and in particular for research on microwave antennae.

Further awards of the Royal Society for 1984 include: the **Copley Medal** to **Professor Subrahmanyan Chandrasekhar**, Morton D. Hull Distinguished Service Professor at the University of Chicago, USA, for his work on theoretical physics, including stellar structure, theory of radiation, hydrodynamic stability and relativity; the **Rumford Medal** to **Professor H.H. Hopkins**, Professor of Applied Optics at the University of Reading, for his many contributions to the theory and design of optical instruments; the **Davy Medal** to **Sir Sam Edwards**, John Humphrey Plummer Professor of Physics at the University of Cambridge for his contributions to the theoretical basis of thermodynamic aspects of polymer chemistry; the **Darwin Medal** to **Professor E. Mayr**, Agassiz Professor Emeritus of the Museum of Comparative Zoology at Harvard University, USA, for his contribution to evolutionary biology; the **Hughes Medal** to **Professor R.P. Kerr**, Professor of Mathematics at the University of Canterbury, New Zealand, for his work on relativity, especially the discovery of the so-called Kerr black hole; the **Leverhulme Medal** to **Professor J.F. Davidson**, Shell Professor of Chemical Engineering at the University of Cambridge, for his contribution to chemical engineering, in particular the use of fluidized beds.

The Royal Society **Wellcome Foundation Prize** has been jointly awarded to **Professor E.R. Andrew**, Professor of Physics at the University of Florida, USA (formerly of the University of Nottingham); **Dr J.M.S. Hutchison**, lecturer in biomedical physics at the University of Aberdeen; **Professor J.R. Mallard** Professor of Biomedical Physics at the University of Aberdeen; and **Professor P. Mansfield**, Professor of Physics at the University of Nottingham, in recognition of their development of NMR imaging as a diagnostic tool in medicine.

At this year's "Analytica" in Munich, the Boehringer prize for Analytical Biochemistry was presented to **Dr Edwin**

Southern for his work on DNA hybridization and its applications in the diagnosis of hereditary diseases.

Lister Institute Research Fellowships have been awarded to **Dr Jeffrey Almond** of the University of Leicester, Department of Microbiology; **Dr Malcolm McCrae** of the University of Warwick, Department of Biological Sciences; **Dr Brian Nunn** of the University of Cambridge, Physiological Laboratory; **Dr Claude Wishik** of the University of Cambridge, Laboratory of Molecular Biology and Department of Psychiatry; and **Dr Stephen Yeaman** of the University of Newcastle-upon-Tyne, Department of Biochemistry.

Applications are invited for the **Andre Lichtwitz** prize awarded by the Institut National de la Santé et de la Recherche Médicale for outstanding clinical or fundamental research on calcium and phosphorus metabolism carried out during the past year by a French or foreign scientist (or team). Conditions of entry for the 1984 prize of 11,000 French francs are obtainable from I.N.S.E.R.M. (c/o Mlle. C. Chirol, 101 rue de Tolbiac 75654, Paris Cédex 13, France). Closing date 30 September 1984.

The **International Union Against Cancer** is looking for candidates for four fellowship programmes (The American Cancer Society-Eleanor Roosevelt-International Cancer Fellowships; the Yamagiwa-Yshida Memorial International Cancer Study Grants; the International Cancer Research Technology Transfer Awards; and the Cancer Research Campaign (UK) International Fellowships). Additional information and application forms may be obtained from the International Union Against Cancer, 3 Rue du Conseil-General, 1205 Geneva, Switzerland.

The **Institute of Metallurgists** is again organizing an International Metallography Competition to encourage the attainment of standards of excellence in the micrographic study of materials. The competition is open to members and non-members of the Institution alike, and this year the total prize money exceeds £1,000. For further details contact Mr B.D. Gibson, Secretary, The Institution of Metallurgists, PO Box 471, 1 Carlton House Terrace, London SW1Y 5BE, UK.

Appointments

Gerard L.E. Turner, DSc, FSA, of the Museum of the History of Science at the University of Oxford, has been elected President of the British Society for the History of Science.

Professor R.J. Elliott, FRS, Wykenham Professor of Physics at the University of Oxford, has been elected Physical Secretary of the Royal Society.

Charles H. Holland, Professor of Geology and Mineralogy at Trinity College, Dublin, has been elected President of the Geological Society, London.

Dr John C. Browne has been appointed to the newly established Chair of Astrophysics at the University of Glasgow.

Sir David Montgomery has been re-appointed for a further five years as Chairman of the Forestry Commission.

Dr Michel Chrétien has been appointed Scientific Director of the Clinical Research Institute of Montreal.

Professor Dr Thomas H. Lee of the Massachusetts Institute of Technology, USA, has been appointed Director of the International Institute for Applied Systems Analysis.

Ms Renate Siebrasse has been appointed Manager, Operations, for Battelle Institute Limited.

Mr Alan Schriesheim has been appointed Director of the Argonne National Laboratory, Illinois, USA.

Dr H.H. Atkinson, Director of Science at the Science and Engineering Research Council, UK, has been elected as the next Chairman of the Council of the European Space Agency, succeeding Professor H. Curiot (France).

Professor G.S.C. Beveridge, Head of the Department of Chemical and Process Engineering at the University of Strathclyde, has been elected President of the Institute of Chemical Engineers.

Dr Roland Schmitt and **Dr Charles Hess** have been elected Chairman and Vice Chairman, respectively, of the National Science Board, of the National Science Foundation, USA.

Dr John R. Durrant, President of the Fox Chase Cancer Centre in Philadelphia, has been designated president-elect of the American Society of Clinical Oncology.

Jim MacNeill, Director of Environment for the OECD, has been appointed Secretary-General of the new World Commission on Environment and Development. Previously he was Commissioner General for HABITAT, the UN Conference on Human Settlements in Vancouver.

Short Courses

24-28 September, **Heat Recovery Network Design**, University of Manchester, (The Registrar, UMIST, PO Box 88, Manchester M60 1QD, UK).

25-28 September, **Capillary Gas Chromatography Course** (Dr J.D. Baty, Packard Instruments Ltd, 13-17 Church Road, Caversham, Berks RG4 7AA, UK).

20 January-1 February, 1985, **Molecular and Cellular Aspects of Immunology**, Rehovot, Israel (Prof. Edna Mozes, Department of Chemical Immunology, the Weizmann Institute of Science, Rehovot, 76100 Israel).

Is all science vocational?

Richard Pearson*

Figures for the numbers of graduates finding employment suggest that it is often difficult to predict the usefulness of a so-called vocational course.

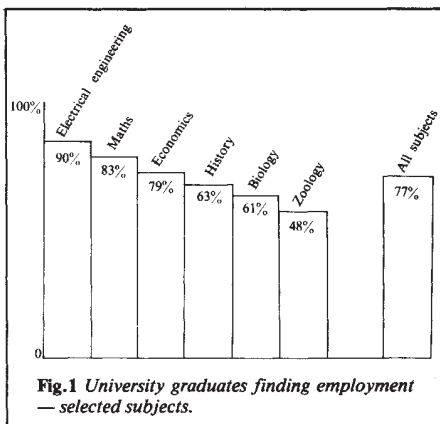
THE pressure is on for a further, more decisive change in higher education policy in the United Kingdom — away from non-vocational towards more vocational courses. But what is vocational? To many, science and technology are vocational while arts and social science, with the notable exception of business-related studies, are not. Life is not, of course, that simple and a new report¹ from the Department of Employment and the Department of Education and Science shows some of the extreme variability between the employability of graduates from different but often seemingly related disciplines. The report is intended to "help intending students form a more realistic assessment of the likely consequences of particular courses of study". It shows subsequent employment patterns of graduates in 1982 and also looks at entry standards and competition for places on degree courses. The poor employment record of many biological science graduates is significant.

Of course, studying for a degree is not just about getting a job and short-term employment prospects. The social aspects of college life are important, while a degree allows the development of academic interests and further academic challenges. Indeed, few graduates cease learning on leaving higher education. Many follow employer-based training. Others stay on for postgraduate study, research or training, an important pre-condition for entry to many jobs such as teaching, which requires a postgraduate diploma, or research, where a PhD is often a pre-condition for entry, although it rarely guarantees employment in itself. More than a third of chemists from the universities have gone on to research after graduation, as have slightly smaller proportions of zoologists, physicists, biologists and geologists. These are far higher proportions than for most subjects and in the case of women graduating in these subjects, as many as ten per cent more have also gone into teacher training. The data for the polytechnics showed rather fewer staying on for further academic study but with broadly similar patterns.

Of those seeking jobs on graduation, it

In last month's Employment article (14 June, p.654), the labels for "Industry" and "Research Institute" in Fig.1 were transposed.

has been those in medicine, engineering, education and law and business-related subjects that have found it easiest to get employment, and significant proportions of these have gone into work related to their degree subject. Perhaps the only surprising group here are the civil engineers, 87 per cent of whom found employment within six months of graduation despite the slump in the civil engineering industry in the United



Kingdom. In part this is explained by significant numbers finding jobs overseas but probably also because they have strong mathematical skills, a major asset for any graduate job-seeker. By contrast, as might be expected. Those finding it hardest to gain jobs have arts backgrounds, where there are few subject-relevant jobs outside teaching; those in the biological sciences also suffer, however; zoologists fared worst in 1982, under half (48 per cent) of those entering the labour market finding permanent employment, a proportion only slightly exceeded by biologists (61 per cent), chemists (66 per cent) and biochemists (65 per cent). The average for all subjects was 77 per cent (Fig.1). Of those finding jobs, only a few scientists directly used their degree subjects. Employment patterns of polytechnic graduates in these subjects were broadly similar but with fewer going quickly into permanent employment. Within each subject, women had slightly better prospects of obtaining early employment than men. The report comments that this may be because a smaller proportion seek to enter the job market, or it may be that women are more flexible in the sort of job they are prepared to do.

For many years, there has been national pressure to increase the number of places in

the vocational areas, if need be at the expense of the non-vocational disciplines. To what extent has the recent reshaping of higher education moved in this direction? Information is now available about admissions to universities in 1983, two years after the first cuts. From these data, the numbers graduating, by discipline, can be estimated for the years to 1986. The total decline in university graduate numbers since 1983 is likely to be of the order of 6,000, if the special cases of medicine and education are excluded. This represents a fall of almost 10 per cent from the peak year for university graduations.

The "lost" graduates have come from across the subject range, particularly arts (33 per cent) and social sciences (34 per cent), but also engineering (14 per cent) and the sciences (16 per cent). On a proportionate basis, the sciences suffered least, a reduction of under 6 per cent, while engineering lost nearly as many graduates as did the arts and social sciences. Within subjects, the contrasts have been even greater, with graduate numbers in electrical engineering (the subject most in demand by employers) falling by 10 per cent, zoology by nearly 30 per cent, chemistry by 4 per cent, and both computer sciences and mathematics each falling slightly over this period. By contrast, biology and biochemistry graduates, both likely to have poor employment prospects, will continue to increase, as will geologists and physicists.

The University Grants Committee has commented that "the reduction in engineering, mathematics and computer science was not unexpected", although it had been hoped to engineer a more decisive switch towards these subjects. Similar data are not, unfortunately, available for the polytechnics, but a similar pattern may emerge. Remedial policies are now in hand to improve the throughput of graduates in information technology and plans may be afoot to shift the balance of higher education further towards the vocational areas. However, given the inertia of institutions and the lack of student interest, something more dramatic may be needed if real flexibility is to be introduced in matching subjects to employability. □

1. *Graduates and Job* (HMSO, London, 1984).

*Richard Pearson is at the Institute of Manpower Studies, Mantell Building, University of Sussex, Brighton BN1 9RF, UK.